Modeling Solar Energy Data Using Periodic Regression

Emily King*, Sango Otieno**, Charles R. Standridge***‡

* Department of Health and Human Services, State of Michigan, Lansing, Michigan

** Statistics Department, Grand Valley State University, Allendale, Michigan

*** School of Engineering, Grand Valley State University, Grand Rapids, Michigan

(KingE7@michigan.gov, otienos@gvsu.edu, standric@gvsu.edu)

‡
Charles Standridge, Grand Valley State University, School of Engineering, Grand Rapids, MI 49546 Tel: +01 616 331 6759, Fax: +01 616 331 6770, standric@gvsu.edu

Received: 01.07.2020 Accepted:16.08.2020

Abstract- Solar energy data was collected from two sites in western Michigan, USA that are not homogeneous with respect to location and solar panels used. Data from one site was collected from August 1, 2009 through July 31, 2019 and the second site from July 12, 2017 through July 31, 2019 and summarized statistically. The average monthly solar energy was higher during the summer and lower during the winter. The variation is higher during the winter and lower during the summer. The annual cycle of average monthly solar energy, as well as the variation, was modeled using periodic regression equations comprised of intercept, sine, and cosine terms as well as an additive term for the difference between the two sites. Model parameters were estimated from all collected data as well as from only the 2017 through 2019 data. All estimated values are statistically significant and consistent in magnitude and sign between the two energy equations. The adjusted coefficients of estimates exceeded 80%. It was concluded that the average monthly solar energy pattern at each site was the same but with different magnitudes and not changing in time. Thus, the model could be applied to all sites in the west Michigan area, now and in the future. For the variation in average monthly solar energy models, the adjusted coefficient of estimation was slightly above 75%. While all parameter values were statistically significant, they were different in magnitude and sign indicating the possibility of a change in variation over time.

Keywords- Solar energy, periodic regression, multiple sites.

1. Introduction

Solar radiation is a primary renewable energy source. The radiation is often harvested using photovoltaic (PV) panels to generate electrical energy. The amount of solar radiation and thus electrical energy varies with the month of the year with more available in the summer months and less in the winter months. Furthermore, the day-to-day variation is higher in the winter months and less in the summer months. Thus, the design of a collection of solar panels, a solar garden, depends on understanding the annual cycle of available solar energy. Mathematical models relating solar energy, both quantity and variation, to the time of the year are helpful in this regard. In general terms, such models are of the periodic type that is a curve that relates a variable (solar energy or variation) to time (month) and is repeated at fixed time intervals[1]. Furthermore, such models are descriptive and are used to help reveal relationships in the existing data to support decision making[2].

Periodic regression is one approach for constructing such mathematical models. The periodic regression method and its usefulness in climatology applications, such as solar radiation, are discussed by Bliss[3]. Čobanović, Lozanov-Crvenković and Nikolić-Đorić[4] describe the periodic regression method as well as including an application to modeling daily maximum UV levels at the University of Novi Sad.

The periodic nature of physical phenomena is modeled using the sine and cosine trigonometric functions. The parameters of these functions are the months of the year expressed as radians. Fig. 1 shows the periodic behavior of the sine and cosine function as a function of the month of the year. Thus, in the periodic regression model, solar energy is the dependent variable and the sine and cosine of the month of the year are independent variables.

In this study, data was collected from solar photovoltaic (PV) panels installed at two sites:

1. Atop the Keller Engineering Laboratories Building (KEB) of Grand Valley State University in Grand Rapids, Michigan, USA from August 1, 2009

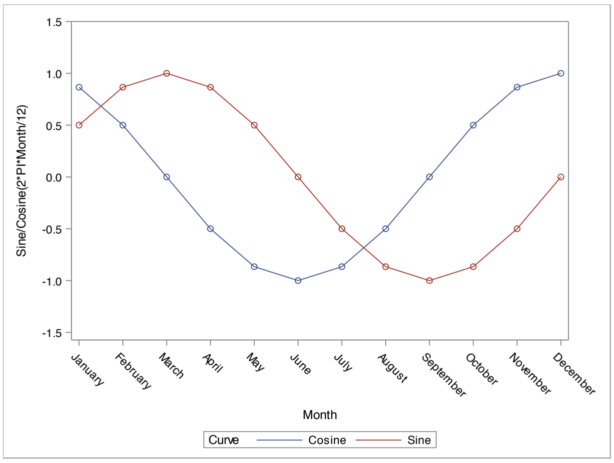


Figure 1. Sine and cosine function by month of the year.

through July 31, 2019. KEB is located at coordinates (latitude, longitude) = (42.9637, -85.6700).

2. Atop a trailer parked at the Hastings Township Office (HTO), Hastings Township, Michigan, USA from July 12, 2017 through July 31, 2019. HTO is located at coordinates (latitude, longitude) = (42.6292, -85.2428).

Note that data were collected at both sites from July 12, 2017 through July 31, 2019, which will be referred to as the overlapping period.

The two sites are not homogeneous. They are 51km apart. At HTO, the solar power generation system consists of eight SS250x type Sonali-manufactured PV solar panels mounted over the top of a semi-truck platform, at the angle that maximizes exposure to sunlight. The rating of this PV system is 250W, 31.32V,7.98A[5]. At KEB, there are 12 solar PV panels manufactured by United Solar Ovonic LLC, model ES 124 with ratings of 124W, 30V, and 4.1A. They are mounted on the roof of the building at a fixed angle of inclination of approximately 44 degrees.

Periodic regression models for solar energy encompassing both KEB and HTO are presented. Model parameter estimation is described and results are discussed. Conclusions are given. First, previous work concerning modeling solar energy, particularly using regression, is reviewed.

2. Background

Yesilbudak, Colak and Bayindir[2] provide a discussion and comprehensive review of solar irradiance and solar power modeling with emphasis on forecasting models. Four different time horizons and their purposes used for solar forecasting are defined and described. The review found that most solar power forecasting models are applied to the first three time horizons of less than one day: less than 15 minutes, 15 minutes to 1 hour and 1 hour to 1 day for purposes including load balancing, reserve capacity planning, and load following.

By contrast, this work concerns descriptive modeling with an annual time horizon based on data from multiple years. Thus, the following literature review emphasizes descriptive models where time is an independent variable while also providing a sampling of work describing various types of models used for both descriptive modeling and forecasting.

Patniak[6] presents a time-based descriptive analysis of power generation at a utility scale solar garden. For individual days, the ideal or maximum power generated by the facility is compared to the actual power generated minute by minute. Dips, significant deviations downward from the ideal power generation, are collected and analyzed for frequency of occurrence, maximum deviation, and average deviation.

Many studies have applied regression techniques for descriptive modeling of solar energy generation and related quantities. Typical of these, Goia and Gustavsen[7] use linear regression to model the energy generated by a rooftop solar plant. Descriptive models are constructed independently for each of six months with the coefficient of determination values, the percent of variation in the data explained by the model, ranging from 80% to 90%. The analysis provides an overview of system performance to aid in understanding the power generated given that the inverters were undersized.

Ianetz et al.[8] compare the potential for solar energy generation at three different sites by statistical analysis of solar radiation data. Distribution function fitting is employed along with computation of monthly and annual averages and coefficients of variation. This descriptive modeling was used to understand and compare the solar irradiation among the three sites.

Kicsiny[9] uses multiple linear regression descriptive models to express the temperature of the fluid leaving a solar collector as a function of the physical properties of the collector, the solar irradiance on the collector surface, and time. These models are shown to perform better than an ordinary differential equation models based in physics principles with respect to predicting the temperature of the fluid leaving the collector.

Kumar, Nagabushanam, and Jayakumar[10] present a time-series approach based in moving averages (MA) to model power generation from a 95kW solar PV plant installed at the Karunya Institute of Technology and Sciences. Data were collected at 15 minutes intervals for 10 days. The models were successfully employed to forecast power generation for one day in the future.

Belmahdi, Louzazni, and El Bouardi[11] discuss the development of models to forecast monthly mean daily global solar radiation for up to three months in the future. Autoregressive moving average (ARMA) and autoregressive integrated moving average models (ARIMA) are used. The latter provide the best results for forecasting based on the AIC and BIC goodness of fit criteria.

Dahmani et al.[12] present the use of an artificial neural network model to predict the solar radiation on a tilted surface from only the measured horizontal global radiation. This is significant since solar panels are almost always installed tilted but only horizontal radiation is usually measured. The model was deemed valid and useful as the root mean square error is 8.81%.

Many studies have used linear regression analysis to predict solar irradiance as a function of astrological and metrological quantities. Typical of these is the study by Paulescu and Blaga[13]. Eight new models are proposed and validated. In the same vein, Keshtegar et al.[14] compare regression models for solar radiation using four different methods. Finally, Li et al.[15] used regression for predicting hourly global solar radiation on a horizontal surface.

Recent work has sought to improve short term forecasting. Ueshima et al.[16] report on a method for improving weekly forecasts of solar radiation statistically from a numerical weather prediction model by considering historical forecasting error occurrences. Colak, Yesilbudak and Bayindir[17] show that a grey wolf optimizer-based multilayer perceptron model is appropriate to predict efficiently daily total horizontal solar radiation. Air temperature, relative humidity and diffuse horizontal solar radiation parameters are evaluated. Al-Hajj, Assi, and Fouad[18] create a one-day solar radiation forecast using a weighted average of a neural network model and a support vector regressors model. Using one year of meteorological data, the combined model was shown to produce superior results versus either model alone.

There is recent work on descriptive modeling as well. Yesilbudak et al.[19] use three different curve fitting methods: Fourier, sum of sines and smoothing spline to model global solar radiation and air temperature parameters at 10-min intervals over a month. Based on the coefficient of determination and the root mean squared error, the smoothing spline model produced the best results. Bosman and Darling[20] demonstrate the need for better estimation of the influence of snow on solar panels with respect to energy generation. In addition, these authors propose a new method for better understanding the return on investment for solar energy systems located in snowy environments.

However, it appears that none of the published studies have used the periodic regression technique to develop descriptive models of the annual cycle of solar energy generation by a solar PV panel system.

3. Methods

First modeling with periodic regression will be discussed. Then additional statistical methods used to analyze the solar energy data will be presented.

The general periodic regression equation is given in equation 1. An additional term to model location is added in equation 2.

$$Y = B_0 + B_1*\text{sine}(2\pi t) + B_2*\text{cosine}(2\pi t) + \text{Error} \quad (1)$$

$$Y = B_0 + B_1*\text{sine}(2\pi t) + B_2*\text{cosine}(2\pi t) + B_3*I + \text{Error} \quad (2)$$

The symbols used in these equations are defined in Table 1.

Table 1. Symbol Definitions

Symbol	Definition
Y	Dependent variable: quantity of or variation in solar energy
T	Time parameter: month / 12
I	Indicator variable representing location: 0 for KEB 1 for HTO
B_0	y intercept
B_1, B_2	Coefficients of the sine and cosine terms
B_3	Coefficient of the location indicator variable

The time parameter refers to a month of the year. For example, July is the seventh month of the year so the time parameter for July is 7/12. The sine and cosine terms work together to model the periodicity. The last term indicates that the periodic behavior does not depend on location. The difference in monthly solar energy or variation in month solar energy is in the difference in magnitude, B_3.

Methods for estimating the parameter values given in equations 1 and 2 are described by Bloomfield[21] as is the more general form of equation 1.

The amount of solar energy generated at each site is observed and reported each day by automated equipment. All observations with a value of zero are considered missing and removed from the data set before analysis. For the KEB data set, there were 170 such values, 4.7%, and for the HTO data set 4 values, 0.55%. For each month, the average, standard deviation and coefficient of variation (CV) are computed from the daily observations. The coefficient of variation is the ratio: standard deviation / average. In addition, the average over all years for each month, January through December is computed for each location.

The coefficient of determination R2 is used as the primary measure of the goodness-of-fit of the periodic regression equations. It is the percent of variation in the data that is explained by a regression equation and thus ranges from 0 to 1. The closer the value of R2 is to 1 the better the fit of the regression equation to the data.

In addition, multiple plots are used to aid in the assessment of model goodness of fit:

1. The monthly solar energy averages (kWh) or CV's calculated from the collected data are plotted against those generated by the regression equation. A good fit is indicated if this plot is approximately a 45-degree line, which occurs when each calculated value is approximately the same as the corresponding observed value.

2. The difference between the calculated monthly solar energy average or CV and the corresponding value generated by the regression equation is known as the residual. Regression analysis assumes that the

residuals are normally distributed with mean 0 and a constant variance.

a. A plot of the residuals versus the values predicted by the regression equation should show random scatter around 0 with width about the same above and below 0.

b. The normal probability plot of the residuals should be a straight line.

c. The histogram of the residuals should be approximately bell-shaped.

The normal probability plot is a type of quantile-quantile or Q-Q plot. Percentile is another word for quantile. The residual values are sorted from smallest to largest and plotted on the y-axis. The corresponding x-axis value for the ith residual in sorted order is the percent point of the standard normal distribution given by the value $P(Z \le i/n)$, where n is the total number of residuals. If the residuals are normally distributed, the plot should show a straight line. Stine[22] provides a brief explanation of the normal probability plot.

Model parameters are estimated and the goodness of fit measures are computed using SAS, the REG procedure.

4. Results

The monthly averages and coefficients of variations of the solar energy generated for the overlapping period are shown by bar graph in Fig. 2 and Fig. 3.

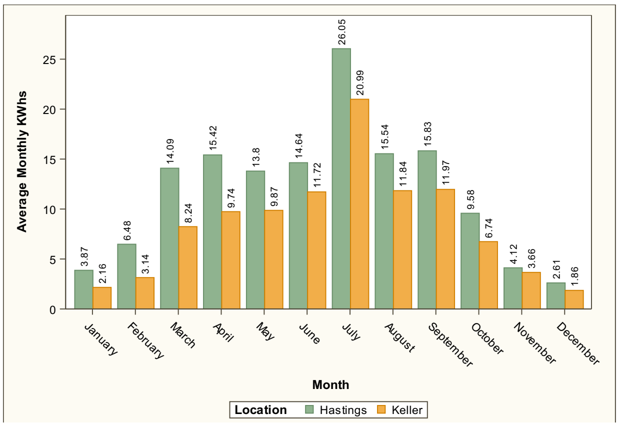


Figure 2. Monthly average energy (kWh) generation at HTO and KEB for the overlapping period.

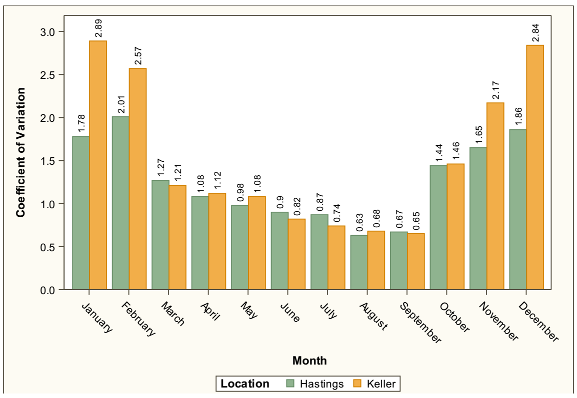


Figure 3. Monthly coefficients of variation at HTO and KEB for the overlapping period.

The parameters of the regression model were estimated from the monthly averages for KEB and HTO computed using all available data, 24 observations total. The results are shown in equation 3 with parameter values expressed to three significant digits, the precision of the data.

$$\text{Predicted Average Energy (kWh)} = 4.01 - 0.838 * \text{sine}(2\pi t) - 2.86 * \text{cosine}(2\pi t) + 1.59 * I \quad (3)$$

All regression parameter values are statistically significant ($\alpha = 0.0135$, at most) as shown in Table 2. Note that the P-value is the probability that the true parameter value is zero. The value of α is the chosen acceptable probability that the true parameter value is zero, which is typically 5%. In this case, the P-values can be interpreted as consistent with choosing an α value as low as 0.0135.

Table 2. Analysis of parameter estimates – all data average monthly energy (kWh).

Variable	Estimate	Standard Error	Test Statistic	P-value	95% Confidence Limits	
Intercept (b_0)	4.01	0.310	13.0	<.0001	3.36	4.65
Sine term (b_1)	-0.838	0.310	-2.71	0.0135	-1.48	-0.192
Cosine term (b_2)	-2.86	0.310	-9.24	<.0001	-3.51	-2.21
Location (b_3)	1.59	0.438	3.62	0.0017	0.673	2.50

The results of the periodic regression are shown graphically in Fig. 4 with the monthly averages displayed and the regression equation superimposed by location.

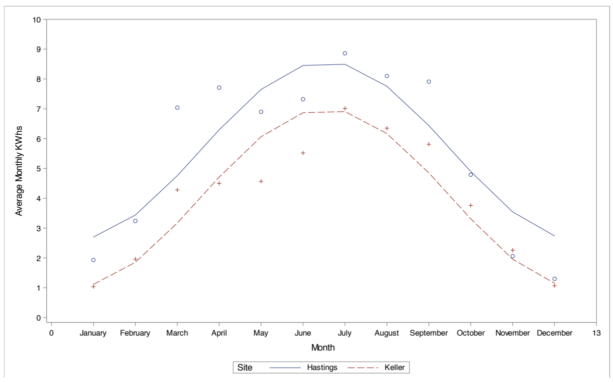


Figure 4. Periodic regression results – all data average monthly energy (kWh).

Alternatively, the parameters of the periodic regression model can be estimated from the data from the overlapping period only, 50 monthly data points. Results are shown in equation 4.

Predicted Average Energy (kWh) =

$$3.96-0.747*\text{sine}(2\pi t)-2.97*\text{cosine}(2\pi t)+1.60*I \quad (4)$$

All regression parameter values are statistically significant ($\alpha = 0.0008$, at most) as shown in Table 3.

Table 3. Analysis of parameter estimates – overlapping period average monthly energy (kWh).

Variable	Estimate	Standard Error	Test Statistics	P-value	95% Confidence Limits	
Intercept (b_0)	3.96	0.205	19.3	<.0001	3.54	4.37
Sine term (b_1)	-0.747	0.208	-3.60	0.0008	-1.16	-0.329
Cosine term (b_2)	-2.97	0.204	-14.6	<.0001	-3.38	-2.56
Location (b_3)	1.60	0.290	5.52	<.0001	1.02	2.19

The results of the periodic regression are shown graphically in Figure 5 with the 25-monthly averages displayed and the regression equation superimposed by location.

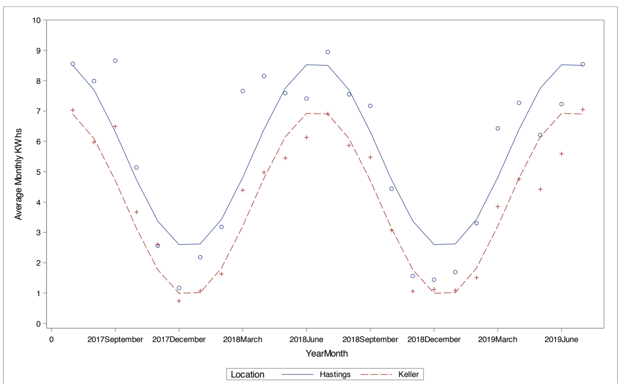


Figure 5. Periodic regression results – overlapping period average monthly energy (kWh).

The adjusted coefficient of determination R2 for the model estimated with all data is 81.7% and for the model estimated with data from the overlapping period is 84.0%. The goodness of fit plots regarding the average monthly energy using all data are shown in Fig. 6 and for the overlapping period data in Fig. 7. The plots are as follows moving clockwise from the upper left.

1. Monthly energy averages (kWh) versus those generated by the regression equation.
2. Residuals versus predicted values.
3. Histogram of the residuals.
4. Normal probability plot of the residuals.

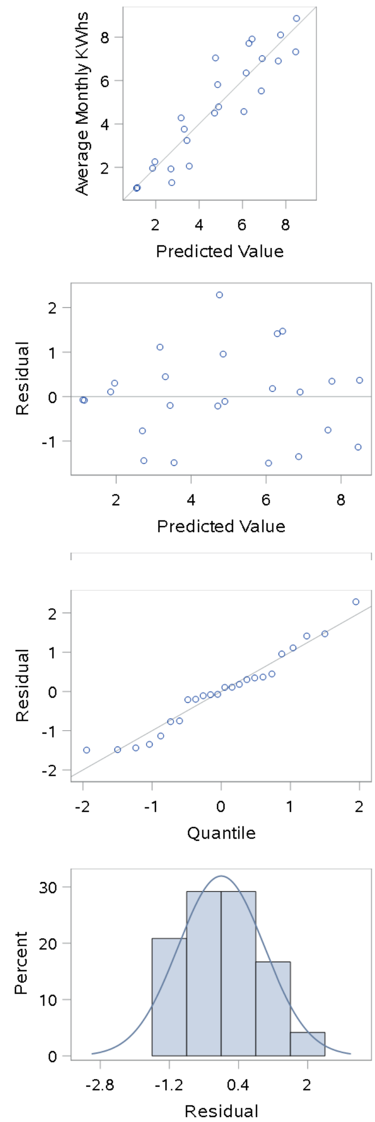


Figure 6. Goodness-of-fit diagnostic plots -- all data average monthly energy (kWh)

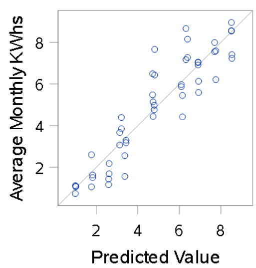


Figure 7. Goodness-of-fit diagnostic plots -- overlapping period average monthly energy (kWh)

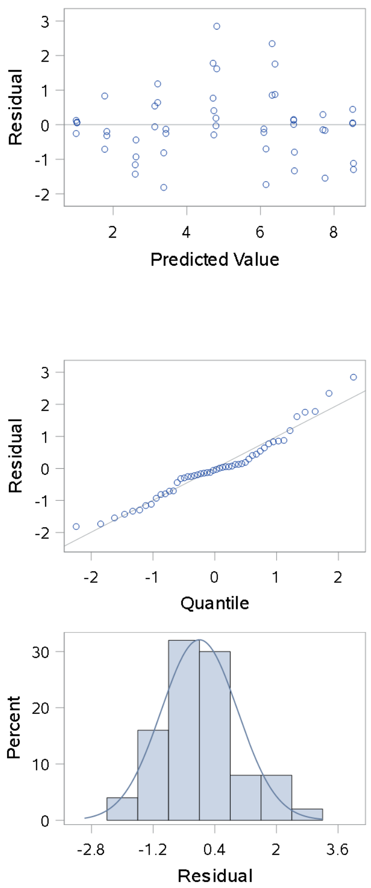


Figure 7. Concluded

The periodicity in the coefficient of variation (CV) can be modeled using the same approach. The parameters of equation 5 are estimated using all data and the parameters of equation 6 are estimated using the data from the overlapping period.

$$\text{Predicted CV} = 0.238 + 0.0825*\text{sine}(2\pi t) + 0.163*\text{cosine}(2\pi t) + 0.192*I \quad (5)$$

$$\text{Predicted CV} = 0.747 + 0.177*\text{sine}(2\pi t) + 0.399*\text{cosine}(2\pi t) - 0.123*I \quad (6)$$

All regression parameter values are statistically significant ($\alpha = 0.0201$, at most) as shown in Tables 4 and 5.

Table 4. Analysis of parameter estimates – coefficients of variation (CV) computed from all data

Variable	Estimate	Standard Error	Test Statistics	P-value	95% Confidence Limits	
Intercept (b_0)	0.237	0.0255	9.33	<.0001	0.184	0.290
Sine term (b_1)	0.0825	0.0254	3.24	0.0041	0.029	0.135
Cosine term (b_2)	0.163	0.0255	6.40	<.0001	0.110	0.216
Location (b_3)	0.192	0.0360	5.34	<.0001	0.117	0.267

Table 5. Analysis of parameter estimates – overlapping period coefficients of variation (CV)

Variable	Estimate	Standard Error	Test Statistic	P-value	95% Confidence Limits	
Intercept (b_0)	0.747	0.0363	20.6	<.0001	0.674	0.820
Sine term (b_1)	0.177	0.0367	4.84	<.0001	0.104	0.251
Cosine term (b_2)	0.399	0.0360	11.1	<.0001	0.327	0.472
Location (b_3)	-0.123	0.0513	-2.41	0.0201	-0.227	-0.020

The results of the periodic regression are shown graphically in Fig. 8 and Fig. 9 with the monthly coefficients of variation (CV) displayed and the regression equations superimposed by location. The adjusted coefficient of determination R2 for the model estimated with all data is 77.0% and for the model estimated with data from the overlapping period is 75.8%. The goodness of fit plots regarding the average monthly energy using all data are shown in Fig. 10 and for the overlapping period data in Fig. 11.

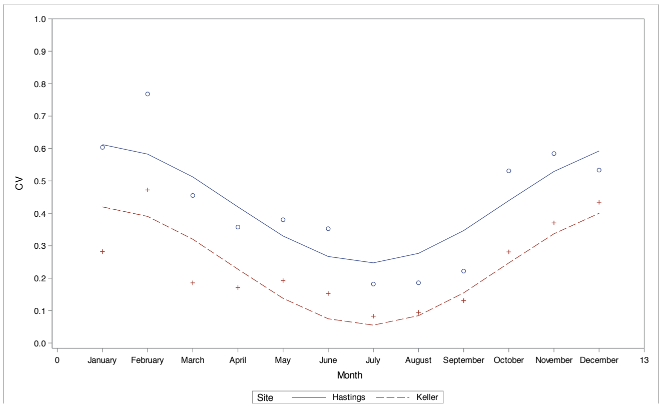


Figure 8. Periodic regression results – coefficients of variation(CV) computed from all data.

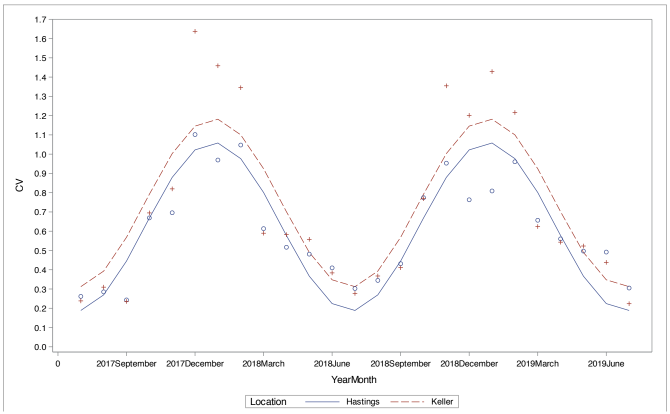
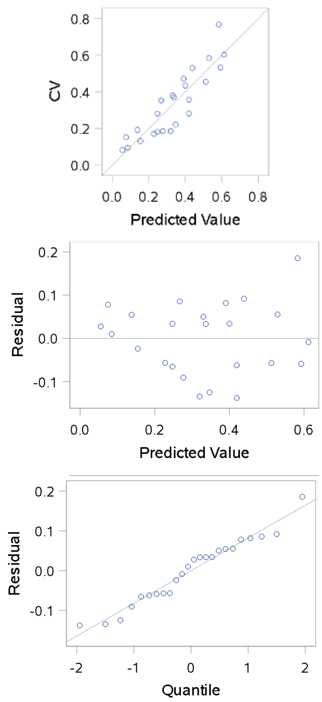


Figure 9. Periodic regression results – coefficients of variation (CV) for the overlapping period.



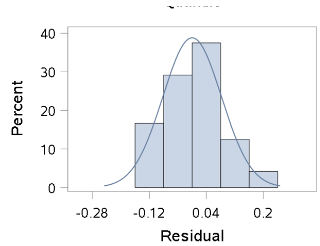


Figure 10. Goodness-of-fit diagnostic plots -- all data coefficient of variation

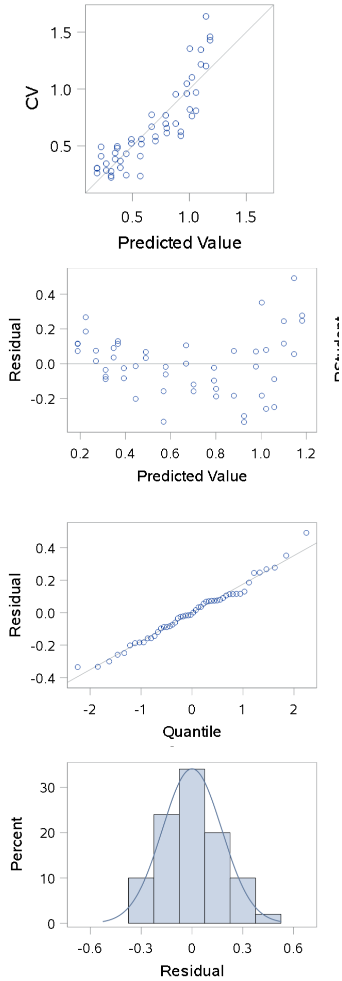


Figure 11. Goodness-of-fit diagnostic plots -- overlapping period for the coefficient of variation

5. Conclusion

Findings have to do with monthly average solar energy generation at HTO and KEB as well as the variation in the monthly average. The annual cycle as well as the magnitude are discussed. Potential changes in each over time are addressed.

Basic statistical summarization of the solar energy data collected at HTO and KEB shows that the average monthly energy generation is highest in the summer, less in the spring and fall, and lowest in the winter. Variation follows the opposite pattern: highest in the winter and lowest in the summer. The average monthly energy is always higher at

HTO than KEB and the within month variation, as measured by CV, is lower at HTO than KEB.

Periodic regression has been shown to effectively model the annual cycle of solar energy collected by PV panels. All parameter estimates of the periodic regression equations are statistically significant ($\alpha = 0.0135$, at most). The sine and cosine terms added together represent the annual cycle with the location indicator term adjusting for the difference in the magnitude of energy generated between KEB and HTO.

Thus, the average monthly solar energy generated at two sites in the same region can be modeled with the same periodic regression equation with the difference in the magnitude modeled by an additive constant. Since this estimated parameter value is statistically significant ($\alpha = 0.0017$, at most) and positive, it can be concluded that the HTO site generates more energy than the KEB site. However, the annual solar energy cycle at each site is the same. This leads to the important conclusion that the annual cycle in average monthly solar energy at any site in the same region, West Michigan, would be the same.

The adjusted coefficient of determination, R2, is over 80% for each energy model, which is consistent with the R2 values obtained by Goia and Gustavsen[7], 80% to 90%, as well as Čobanović, Lozanov-Crvenković and Nikolić-Đorić[3], 81%. The model with parameter values estimated from all the data is almost identical to the model with parameter values estimated from the overlapping period only. All model parameter values are statistically significant and of the same magnitude and sign. In addition, the largest difference in value is for the sine term, about 11%. Thus, it can be concluded that the annual solar energy cycle is not changing in time. Thus, the model can be used for future planning.

In the same way, the annual cycle of variation can be modeled using periodic regression. All parameter values are statistically significant ($\alpha = 0.0201$, at most). Again, the difference in the magnitude of the variation between KEB and HTO can be modeled using an additive term with a statistically significant coefficient ($\alpha = 0.0201$, at most). This leads to the conclusion that the annual cycle in the variation in average monthly solar energy at any cite in the West Michigan region would be the same.

The adjusted coefficient of variation for each model is slightly above 75%, an acceptable fit for physical phenomena. However, the estimated parameter values for the two models vary noticeably. The magnitudes of coefficients for the intercept, sine, and cosine term are noticeably greater for the overlapping period model. The sign of the location term is different between the two models. This indicates a possible change in average monthly solar energy variation over time.

The graphs in Fig. 6, 7, 10 and 11 reinforce that the models fit the data well. The residuals appear to be normally distributed as seen in the normal probability plots and the histograms. Note that the residuals for the overlapping period models fit tighter to the 45-degree line in the normal probability plot than the residuals for the models with all data indicating a better fit to the data for the former. The residuals randomly scatter on either side of zero within the same distance from zero except for one positive outlier. The plot of the actual data versus the estimates from the regression equation follow a 45-degree line well.

In summary, the primary findings are as follows:

1. Average monthly energy generation is highest in the summer and lowest in the winter with variation following the opposite pattern.
2. Periodic regression effectively models the annual cycle of solar energy, both the monthly average and the variation in the monthly average.
3. The average monthly solar energy generated at two sites in the West Michigan region can be modeled with the same periodic regression equation with the difference in the magnitude modeled by an additive constant. This is also true for the variation in the average monthly solar energy.
4. By inference based on the periodic regression modeling results, the annual cycle in average monthly solar energy at any site in the West Michigan region is the same. This is also true for the annual cycle of the variation in average monthly solar energy.
5. The annual cycle in average monthly solar energy in West Michigan is not changing in time. Thus, the periodic regression model can be used for future planning.
6. A potential change in the magnitude of the variation in average monthly solar energy over time was identified.

References

[1] T. M. Little and F. J. Hills, *Agricultural Experimentation Design and Analysis*. New York: John Wiley and Sons, Inc, 1978.

[2] M. Yesilbudak, M. Colak and R. Bayindir, What are the current status and future prospects in solar irradiance and solar power forecasting? *International Journal of Renewable Energy Research,* vol 8, no. 1, pp. 635-648, 2018.

[3] C. I. Bliss, *Periodic Regression in Biology and Climatology*, New Haven: The Connecticut Agricultural Experiment Station, 1958.

[4] K. Čobanović, Z. Lozanov-Crvenković and E. Nikolić-Đorić, "Periodic regression", ICOTS-7 2006, Savador, Bahia, Brazil. Retrieve April 27, 2020 from: https://iase-web.org/documents/papers/icots7/C207.pdf?1402524966

[5] C. R. Standridge, L. Corneal, and N. Baine, Advances in repurposing and recycling of post-vehicle-application lithium-ion batteries, *Mineta National Transit Research Consortium,* 2016. Retrieved April 28, 2020 from: https://transweb.sjsu.edu/mntrc/research/Advances-Repurposing-and-Recycling-Post-Vehicle-Application-Lithium-Ion-Batteries.

[6] A. Patnaik, A statistical analysis of a 4.5 MW solar energy center, *Masters Theses, 7585*, 2016 Retrieved March 3, 2020 from https://scholarsmine.mst.edu/masters_theses/7585.

[7] F. Goia and A. Gustafsen, "Energy performance assessment of a semi-integrated PV system in a zero emission building through periodic linear regression method", *11th Nordic Symposium on Building Physics, NSB2017*, Trondheim, Norway, 11-14 June 2017.

[8] A. Ianetz, V. Lyubansky, I. Setter, E. G. Evseev and A. I. Kudish, "A method for characterization and inter-comparison of sites with regard to solar energy utilization by statistical analysis of their solar radiation data as performed for three sites in the Israel Negev region", *Solar Energy*, Vol. 69, No. 4, pp. 283–293, 2000.

[9] R. Kicsiny, "Multiple linear regression based model for solar collectors", *Solar Energy*, Vol. 110, pp. 496-506, 2014.

[10] S. S. Kumar, P. Nagabushanam and M.S Jayakumar, "Forecasting the solar power generation by modified time series model for cyber-physical power system", *International Journal of Pure and Applied Mathematics*, Vol. 119, No. 15, pp. 1457-1462, 2018.

[11] B. Belmahdi, M. Louzazni, and A. ElBouardi, "One month-ahead forecasting of mean daily global solar radiation using time series models", *Optik*, Vol. 219, 2020. https://doi.org/10.1016/j.ijleo.2020.165207

[12] K. Dahmani, R. Dizene, G. Notton, C. Paoli, C. Voyant, and M. L. Nivet, "Estimation of 5-min time-step data of tilted solar global irradiation using ANN (Artificial Neural Network) model", *Energy,* Vol. 7, No. 1, pp 374-381, 2014. https://doi.org/10.1016/j.energy.2014.04.011

[13] E. Paulescu and R. Blaga, "Regression models for hourly diffuse solar radiation", *Solar Energy*, Vol. 125, pp. 111-124, 2016. http://dx.doi.org/10.1016/j.solener.2015.11.044.

[14] B. Keshtegara, C. Mert, and O. Kisic, "Comparison of four heuristic regression techniques in solar radiation modeling: Kriging method vs RSM, MARS and M5 model tree", *Renewable and Sustainable Energy Reviews*, Vol. 81, Part 1, pp 330-341, January 2018. DOI: 10.1016/j.rser.2017.07.054

[15] D. H.W.Li, W. Chen, S. Li, and S. Lou, "Estimation of hourly global solar radiation using Multivariate Adaptive Regression Spline (MARS) – A case study of Hong Kong", *Energy*, Vol. 186, November 2019. https://doi.org/10.1016/j.energy.2019.115857.

[16] M. Ueshima, T. Babasaki, K. Yuasa and I. Omura, "Examination of correction method of long-term solar radiation forecasts of numerical weather prediction", *8th International Conference on Renewable Energy Research and Applications (ICRERA)*, Brasov, Romania, 3-6 November 2019, pp. 113-117. doi: 10.1109/ICRERA47325.2019.8997070.

[17] M. Colak, M. Yesilbudak and R. Bayindir, "Forecasting of daily total horizontal solar radiation using grey wolf optimizer and multilayer perceptron algorithms", *8th International Conference on Renewable Energy Research and Applications (ICRERA)*, Brasov, Romania, 3-6 November 2019, pp. 939-942. doi: 10.1109/ICRERA47325.2019.8997040.

[18] R. Al-Hajj, A. Assi and M. M. Fouad, "Forecasting Solar Radiation Strength Using Machine Learning Ensemble," *2018 7th International Conference on Renewable Energy Research and Applications (ICRERA)*, Paris, 14-17 October 2018, pp. 184-188, doi: 10.1109/ICRERA.2018.8567020.

[19] M. Yesilbudak, M. Colak, R. Bayindir and H. I. Bulbul, "Very-short term modeling of global solar radiation and air temperature data using curve fitting methods," *IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA)*, San Diego, 5-8 December 2017, pp. 1144-1148, doi: 10.1109/ICRERA.2017.8191233.

[20] L. Bosman and S. Darling, "Difficulties and recommendations for more accurately predicting the performance of solar energy systems during the snow season", *2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*, Birmingham, 2016, pp. 567-571, doi: 10.1109/ICRERA.2016.7884398.

[21] P. Bloomfield, *Fourier analysis of time series, 2nd ed*, New York: John Wiley & Sons, Inc., 2000.

[22] R. A. Stine, "Explaining normal quantile-quantile plots through animation: the water-filling analogy," *The American Statistician*, vol 71, No. 2, pp. 145-147, 2017. DOI: 10.1080/00031305.2016.1200488