

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/300128195>

# Estimation of missing values in solar radiation data using piecewise interpolation methods: Case study at Penang city

Conference Paper · December 2015

DOI: 10.1063/1.4937079

CITATIONS

2

READS

139

3 authors, including:



**Azizan Saaban**

Universiti Utara Malaysia

120 PUBLICATIONS 580 CITATIONS

[SEE PROFILE](#)



**Mohd Nazari A Bakar**

Universiti Teknologi MARA (Perlis), Perlis, Malaysia

32 PUBLICATIONS 631 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



national university of Malaysia [View project](#)



An ARMA-Markov Chain Based Fuzzy Time Series Model For Forecasting Oil Production And Oil Consumption [View project](#)

## **Estimation of missing values in solar radiation data using piecewise interpolation methods: Case study at Penang city**

Mohd Lutfi Zainudin, Azizan Saaban, and Mohd Nazari Abu Bakar

Citation: [AIP Conference Proceedings](#) **1691**, 040029 (2015); doi: 10.1063/1.4937079

View online: <http://dx.doi.org/10.1063/1.4937079>

View Table of Contents: <http://scitation.aip.org/content/aip/proceeding/aipcp/1691?ver=pdfcov>

Published by the [AIP Publishing](#)

---

### **Articles you may be interested in**

[Evaluation of linear interpolation method for missing value on solar radiation dataset in Perlis](#)

AIP Conf. Proc. **1660**, 050024 (2015); 10.1063/1.4915657

[Estimation of missing rainfall data in Pahang using modified spatial interpolation weighting methods](#)

AIP Conf. Proc. **1643**, 65 (2015); 10.1063/1.4907426

[Estimation of missing rainfall data using spatial interpolation and imputation methods](#)

AIP Conf. Proc. **1643**, 42 (2015); 10.1063/1.4907423

[On piecewise interpolation techniques for estimating solar radiation missing values in Kedah](#)

AIP Conf. Proc. **1635**, 217 (2014); 10.1063/1.4903586

[Extreme Value Modeling and Prediction of Extreme Rainfall: A Case Study of Penang](#)

AIP Conf. Proc. **1309**, 372 (2010); 10.1063/1.3525139

---

# Estimation of Missing Values in Solar Radiation Data using Piecewise Interpolation Methods: Case Study at Penang City

Mohd Lutfi Zainudin<sup>1,2, a)</sup>, Azizan Saaban<sup>1, b)</sup> and Mohd Nazari Abu Bakar<sup>3, c)</sup>

<sup>1</sup>*School of Quantitative Sciences, UUMCAS, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia.*

<sup>2</sup>*Institut Matematik Kejuruteraan (IMK), Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia.*

<sup>3</sup>*Faculty of Applied Science, Universiti Teknologi Mara, 02600 Arau, Perlis, Malaysia.*

<sup>a)</sup>Corresponding author: mdlutfi07@gmail.com

<sup>b)</sup>azizan.s@uum.edu.my

<sup>c)</sup>mohdnazari@perlis.uitm.edu.my

**Abstract.** The solar radiation values have been composed by automatic weather station using the device that namely pyranometer. The device is functions to records all the radiation values that have been dispersed, and these data are very useful for it experimental works and solar device's development. In addition, for modeling and designing on solar radiation system application is needed for complete data observation. Unfortunately, lack for obtained the complete solar radiation data frequently occur due to several technical problems, which mainly contributed by monitoring device. Into encountering this matter, estimation missing values in an effort to substitute absent values with imputed data. This paper aimed to evaluate several piecewise interpolation techniques likes linear, splines, cubic, and nearest neighbor into dealing missing values in hourly solar radiation data. Then, proposed an extendable work into investigating the potential used of cubic Bezier technique and cubic Said-ball method as estimator tools. As result, methods for cubic Bezier and Said-ball perform the best compare to another piecewise imputation technique.

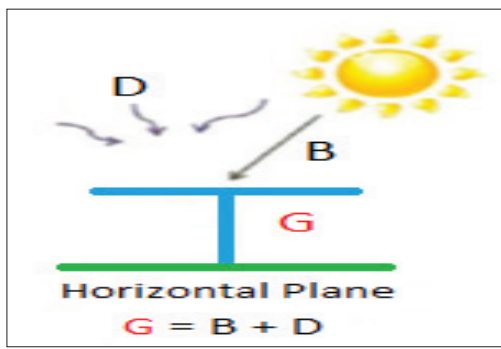
## INTRODUCTION

Monitor the behavior of solar radiation intensity played the vital issue into measuring the solar-energy concentration on the specific area which is very significantly for any solar-energy projects based. On Malaysia's country, the solar radiation intensity has been composed at weather stations due to have pyranometer instrumentation. Unfortunately, only selected weather stations have been equipped with that device. This instrument has been recognized as the main apparatus which measuring the solar radiation that dispersed on the ground surface. In such, developing countries like Malaysia applied that such sophisticated monitoring instrument mostly to obtain the attribute of solar radiation intensity due to abilities that offered such as affordable to measure need's variables at continuously and store the data quantity in largely. On Figure 1(a) have been shown the device used to measure the amount of solar radiations that received from the atmosphere throughout the given times, and the device is namely pyranometer. The pyranometer has defined as the hemisphere sensors that placed on the horizontal position [2] and uses several elements likes thermo-electric, photoelectric, pyro-electric or bimetallic as the sensor [3]. Here, Figure 1(b) has been explained about the pyranometer device which is placed at the horizontal position. Normally, the devices locate at the isolated area such as area where the weather stations are always being developed.

Solar radiation has been identified as the transmitted light from the sun that contents radiation or electromagnetic waves which capable of generate a lot of energy through heat and light liberation. The radiation that contained inside the sunlight is come from the fusion atom which is released from the sun and transform into the electromagnetic waves [4] and introduced as the visible light. Truly, the visible light or radiant lights are formed of solar energy which is revealed the energy when it falls down on the earth's surface. Then, continues measurement is a perfectly situation into observing solar radiation intensity due to given precisely values on long-term data, which extremely used on solar energy-system performance evaluation. Therefore, there is no wonder the acquisition data of solar intensity have

recorded at continuously and have presented at regularly into energy unit such as energy per unit time per unit area or power per unit area [5].

The recorded solar radiation data can be described in terms of time-series data [6]. Time-series data is about the real-world data acquisition which representative the repeated measurement data or continuous data at the certain period of times. The solar radiation time series is directly vital to solar data analysis and as the input variable for another related scientific research. However, having been missing values frequently infects time-series data [7] especially for solar radiation time-series data [3], [8], [9], [10]. In earlier research, [11] has reported the unreliable solar radiation data rise when the absent and malfunction instrumentation occurs and this circumstance needs model estimation to keep the data reliable and accurate. As it knew, solar radiation data is mostly difficult to measure compares to grab other meteorological parameter values due to these problematic, technical failure, operation-related problems and expensive device [2], [12], [9]. Both problems have created mainly contribution that provides lower quality data because of existed missing values, immediately offered the difficulty task for solar radiation data analysis. Reliable, accurate and complete solar radiation data is very significant to modeling, designing, predicting process and inclusive solar radiation system performance evaluation due to availability high-quality data [8], [9], [13], [14]. Without the higher quality and reliability data provides for data analysis, the entire modeling and validating process produced the bias result in the conclusion.



(a)



(b)

**FIGURE 1.** (a) The method applied to allocate the measuring devices (pyranometer) at the horizontal plane to observe global solar radiation that reached on the ground level ( $G$  = Total radiation,  $B$  = Beam radiation,  $D$  = Diffuse radiation) (b) Pyranometer device which used at Weather Station

## Problematic of Missing Values

Generally, missing values in time series can be defined as the formation of absent values on the continuous fluctuate data recording. Missing values are characterised by presence of absent values that appear on the observed data by several factors which mainly comes from the monitoring instrument such as monitoring instrument interruption, malfunction sensor on the machine, and network reorganization [1]. Into dealing acquires sequence's informative values at certain period of time using continuous automatic instrumentation has to confront by some amounts of absent values frequently [7]. Incomplete data is generated when missing values on the data set have been raised and invited to reduce prediction power of efficiency through decrease amount of sample size and raise bias sample [15]. Addition, operational maintenance and faulty data reading are also corresponding causes that offered to missing values attached on the observed data [9]. Therefore, researchers have tried to avoid used incomplete data sample merely to generate reliable values for data analysis.

Missing values occasionally proud to be appeared on the measured continues data when the intelligent monitoring device involves grabbing importance parameter values to store for data analysis and research studies. In terms of solar radiation data, on both, long-term data ( $> 1$  year) and short-term data are vital for data analysis. The long data is significant for solar-energy systems performance determination [16], in while short-term data importance for agriculture, urban planning and atmospheric pollution analysis [17]. However, available measured solar radiation data

is often insufficient due to creation of missing values. The prerequisite condition for solar radiation data analysis either short-term or long-term data is the availability of the complete data before the data is modeled [18]. Without available, the complete data before the data analyses inaccurate results might be generated [19] especially for that research which used run-off modeling process as needs [20]. Thus, missing values is capable of disrupt solar-energy system development such as the design of solar-energy conversion system and solar-energy device efficiency inclusive other related scientific research that used solar radiation data as the input parameter.

Solar radiation is one of the climatology parameters which vital need recently for climatology studies. Today's, climatology studies have drifted to investigate about the world climate changing and global warming situation due to its high reactivity to the anthropogenic interferences [21]. Intended missing values of several climatology parameters brought this matter as the highlighted problematic in climatology research [7]. Regarding to [22] the climatology studies need parameter values that presence without missing value which usually appear inside the data set due to validation process because of the statistical properties of the sequences values able to be preserved. Hence, complete data of solar radiation not only relevant on modeling and solar-energy system evaluation but also on other related studies such as global warming and climatic change investigation. Therefore, a meticulous methodology ought to be responsible to manage missing values that arise in time-series data, and then, the statistical properties on the solar sequence's data able be protected.

A lot of researchers have admitted the existing missing values on solar radiation time series since in memorial until present [3], [8], [9], [10], [11]. In addition, limited numbers of research have been touched and solved this regards problem because fewer awareness among the researcher about the missing values implication towards time series prediction, which extremely apply for modeling and optimum device's design through the data analysis. The sequence of solar radiation data is very significantly for solar data analysis due to the performance prediction on the solar energy system that has developed. Prediction on solar energy system is importance for ensures the reliability through the system use, and the complete time-series data ought to be applied as preconditions before the solar-energy systems performance is evaluated [8]. As solution, researcher [3], [8], [19] agree the missing values that have existed on solar radiation data absolutely to estimate due to provide reliable and high-quality data before the data being analyst and simultaneously produce accurate results in the conclusion.

## **Review Works on Estimation Missing Values on Solar Radiation Data**

Estimations missing values or interpolation works are tended to predict missing values in time-series data, in other words, employ imputation approaches, which desire to replace the missing observation from the sequence's data by plausible imputed values [22]. In terms of technical, missing values imputation is a strategy that has employed to predict missing value in time series using mathematical approaches, which designed from the remains data distribution. The replacement values on time-series data by predicted values have been introduced as missing values imputation and a significant method tended to complete time-series data before the data analysis has been made. Unfortunately, estimation works into dealing missing values on solar radiation data as the main research's objective is very least to propose.

Only a few works have been proposed into dealing missing values on solar radiation data since until the present. On the past studies, mostly work prefers to reconstruct the available data through used available model techniques such as sunshine based model due to avoid unreliable data, which can be produced bias for data analysis. However, today's increasing awareness among researchers about the available data, the quality data used before the model is implemented to rebuild time-series data and raise questionable about the numbers of missing values [23]. Followed by another paper, [24] instantly a barrier for the time-series data prediction scheme existed when missing values occur in the data set because of time series prediction requested continues data as condition. Presently, has a little works into dealing missing values on solar radiation data in an effort to predict absent values and replace with imputed values. Therefore, missing value prediction studies on solar data still at the primitive or vague stage. On this section, review works on lately studies that regard on missing values estimation working on solar radiation data is reported.

The research that gives out about the missing value estimation on hourly solar radiation data, at first has proposed by [8]. This article straightly focuses on estimation works using model based techniques, which relay on the meteorological parameter that used as the input parameter. Two methods have been used as the comparison that consists by available temperature model based and followed by the proposed techniques which have two approaches, which are applied for decision matrix system and regression correlation but in while the temperature based models are put on Hargreaves-Samani model and Hargreaves-Samani-Annandale model. As the result, the proposed method using the decision matrix system had the best performance estimation. However, the proposed estimation model still under

request for additional parameter input for computational process and absolutely contributes highly cost for estimation. Accordingly to the limitation that offered is very nicely to move on another approach method to solve the matter.

On research proposed by [3] also works on missing data imputation of solar radiation data but using ten-minute basis data. The main target of this research is to examine the proposed imputation methods in order to estimate missing or wrong values of received solar data. The methods proposed for this research using multivariate imputation as missing value estimator, then have applied in the difference area. Three multivariate technique's namely Multivariate Imputation by Chained Equations (MICE), Inverse Distance Weighting (IDW) and Multiple Linear Regression (MLR) put into an application for this single study for an evaluation. In the overall analysis, the result that has appeared supported the performance of MICE imputation technique on solar radiation data is better than the remains techniques. The exploration study on the MICE has been exposed that imputation has a bigger potential to become an appropriate imputation technique to predict missing values on solar data. Unfortunately, MICE method needs to spend more times on the computational process and necessary not suitable used for univariate data application, then require additional parameter input. Secondly, the best method that has been reported only valid for minutes data basis but not in hourly data used due to the rare event for MICE application.

Then, have followed by [1] who have been proposed research on missing value imputation that applied on solar radiation data. The proposed investigation is regarded on piecewise interpolation approaches that imply on hourly solar radiation data to substitute the existed missing values by imputed values. The data applied to this research using simulation missing values data set consist with five kinds of missing value percentage. The simulation data that used in the study are 10%, 20%, 30%, 40%, and 50% missing values, which obtained from Alor Setar area throughout year 2012 have been taken from Malaysian Meteorology Department. The piecewise interpolation methods that have been practiced are Nearest Neighbor, Linear, Cubic, Splines, Bezier and Said-Ball interpolation. Piecewise interpolation does not as usual methodology that proposes as missing value predictor, especially on solar radiation data. Bezier and Said-Ball interpolation are techniques that very rare applicant who deserves as missing value predictor and then the potential on both approaches as missing value estimator has reported in this investigation. As the result, has reported the piecewise techniques for Bezier and Said-Ball have been served the best performance effect among other methods and the both methods presently similar output results. The advantages that offered when using interpolation techniques is reliable applied on single data and never attend any additional parameter input for computational process.

The idea to propose this article has triggered based on the [1] dealing works which more focus on the potential ability of cubic Bezier and Said-Ball that offered. On both techniques had to preview similar performance and shown the best performance effect among other piecewise imputation's methods. However, the results appeared to keep doubtful due to the performance methods applied still underneath work's umbrella investigation. The expendable work on missing value imputation towards measured solar radiation data is proceeded and the potential on the cubic Bezier and Said-Ball approaches is reported as overcomes solution. Both are the main respect's issue that has been carried out through this paper. This article is unveiling the potential of piecewise interpolation techniques as missing value predictor on solar radiation data for Penang's city. Penang city has been selected for this investigation due to one of the northern areas, which potentially used solar radiation as new energy sources to generate electricity. As it knew, Penang is an island which fully with the ironic places that can be an interested tourist for visits and much of the industrial sector had an interest to build their factories on this land. Both reasons are the main factor that contributed for high electricity demands and have to fulfill for ensures the economic sector will continue to growth. This situation has encouraged the government to use renewable energy from the sunlight to generate electricity for industrial and residence area. Without reliable and accurate data, the project costing and technologies should be implemented are unpredictable. Then, the brilliant idea for our good deserves to bury.

## DATA SET

Here, hourly solar radiation intensity data is a main parameter that is used in this current study. This data has obtained from the Malaysian Meteorology Department due to avoid invalid data applied. The solar radiation intensity data is measured in Mega Joule per meter square ( $\text{MJm}^{-2}$ ) and observed from 8 am to 7 pm throughout on year 2012 has been chosen for use for this study. Then, the total of data point that involved on this data set is about 3660, which are provided for complete data values. Based on the complete data values have been generated for five formatting of missing value simulation data sets that consists by 10%, 20%, 30%, 40%, and 50%, then employed as an effort to examine the prediction performance towards the proposed methods. The simulation data that obtained from the differences level of missing values for this study is produced using Statistical Package for Social Science (SPSS).



## CUBIC BÉZIER AND CUBIC SAID-BALL

The current research application is referred to develop piecewise interpolation techniques to estimate missing values on the simulation data set. On the available simulation, data set is divided by several parts which each part consists a few numbers of data which the data conditions are  $a \leq x_l < x < x_k \leq b$  and the missing data ( $x$ ) have been included on the intermediate data set. Then the missing data needs to predict using proposed interpolation techniques based piecewise technique. Interpolation technique is defined as a function that is used to predict the value at the specific data point using available interval data points,  $y = F(x)$ .

The development of Cubic Interpolation that contributes for Bézier and Said-Ball approaches usually required four data points which two data points refer as initial data point ( $P_0$ ) and end data point ( $P_3$ ) while the remains data points ( $P_1$  and  $P_2$ ) which located at the intermediate data point,  $P_0$  and data point,  $P_3$ . Both data points,  $P_1$  and  $P_2$  also knew as control data points which responsible to control the curve development. Figure 2 has described the whole environment on cubic Bézier and Said-ball scenario techniques application.

Given interval data point for hourly solar radiation intensity data,  $[a, b]$  where  $a = (x_l, y_l)$  contribute as interval start point and  $b = (x_k, y_k)$  is introduced as the ends of interval data point where  $k$  is hour's numbers in condition  $2 \leq k \leq 8740$ . Thus, when implies the interval data point on cubic Bezier and Said-Ball interpolation,  $P_0 = y_l$  and  $P_3 = y_k$  while the control points ( $P_1$  and  $P_2$ ) ought to be calculate first before the both techniques directly interpolate the missing hourly solar radiation that existed on the interval data point. The intermediate control points  $P_1$  and  $P_2$  can be found through the tangent in the start interval point ( $P_0$ ) and at the last interval point ( $P_3$ ) respectively. Thus,  $P_1$  and  $P_2$  can be determined as following below equation:-

$$P_1 = P_0 + \frac{1}{3}(x_k - x_1)F'(x_1) ; P_2 = P_3 + \frac{1}{3}(x_k - x_1)F'(x_k) \quad (\text{Bezier}) \quad (1)$$

$$P_1 = P_0 + \frac{1}{2}(x_k - x_1)F'(x_1) ; P_2 = P_3 + \frac{1}{2}(x_k - x_1)F'(x_k) \quad (\text{Said-Ball}) \quad (2)$$

where  $F'(x_1)$  and  $F'(x_k)$  are found by estimated gradients at start and ends of the hourly interval. The gradient point means the estimation point that engaged the well-introduce finite difference approaches which are backward, forward, and central difference methods.

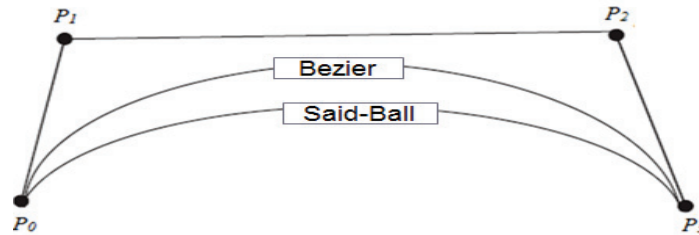


FIGURE 2. The cubic Bézier and cubic Said-Ball curves

On the degree, three Bézier and Said-Ball interpolation methods are normally used in application but very rare to be found on estimation dealing works on solar radiation data, particularly on missing value imputation. Cubic Bézier curve and Said-Ball curve in regularly have been controlled by four data point,  $P_0, P_1, P_2$  and  $P_3$ . On both data points,  $P_0$  and  $P_3$  are known data points and in between them have been existed the data point  $P_1$  and  $P_2$ . Regarding to the figure 1, the curve is initiated at  $P_0$ , and then the straight lines or the tangent lines go against  $P_1$ . At the ends of the interval data,  $P_3$  is always appeared after pass through the tangent lines from  $P_2$ . If both are controlled by four-point  $P_0, P_1, P_2$  and  $P_3$ , then, the curve is respectively given by:-

$$B(t) = (1-t)^3 P_0 + 3t(1-t)^2 P_1 + 3t^2(1-t) P_2 + t^3 P_3 \quad (\text{Bezier}) \quad (3)$$

$$S(t) = (1-t)^2 P_0 + 2t(1-t)^2 P_1 + 2t^2(1-t) P_2 + t^2 P_3 \quad (\text{Said-Ball}) \quad (4)$$

where  $t = \frac{x-x_1}{x_k-x_1}$

According to (3) and (4), both equations have similar properties except for its basis function, which derived from their own basis function. Bezier curve basis function has released from the Bernstein polynomial basis function due to the definition of the Bezier curve itself. A Bezier curve of degree  $n=3$  are defined by:-

$$B(t) = \sum_{i=0}^3 P_i B_{i,3}(t), \quad 0 \leq t \leq 1 \quad (5)$$

where  $P_i$  have recognized as the control points and the  $B_{i,3}(t)$  are the polynomials of degree  $n=3$ , called cubic Bernstein polynomials and have been represented as cubic Bezier basis function. The Bernstein polynomials,  $B_{i,n}(t)$  of degree  $n=3$  are defined by:-

$$B_{i,3}(t) = \binom{3}{i} t^i (1-t)^{3-i}, \quad i = 0, 1, \dots, 3. \quad (6)$$

where  $\binom{3}{i} = \frac{3!}{i!(3-i)!}$

Next, the Said-Ball curve of degree  $n = 3$  are established by:-

$$S(t) = \sum_{i=0}^3 P_i S_{i,3}(t) \quad (7)$$

where  $P_i$  have knew as the control points and the  $S_{i,n}(t)$  is Said-Ball basis function which has clarified below:-

$$S_{i,3}(t) = \begin{cases} (i+1)t^i(1-t)^2, & 0 \leq i \leq 1 \\ S_i^3(1-t), & 2 \leq i \leq 3 \end{cases} \quad (8)$$

## RESULTS AND DISCUSSION

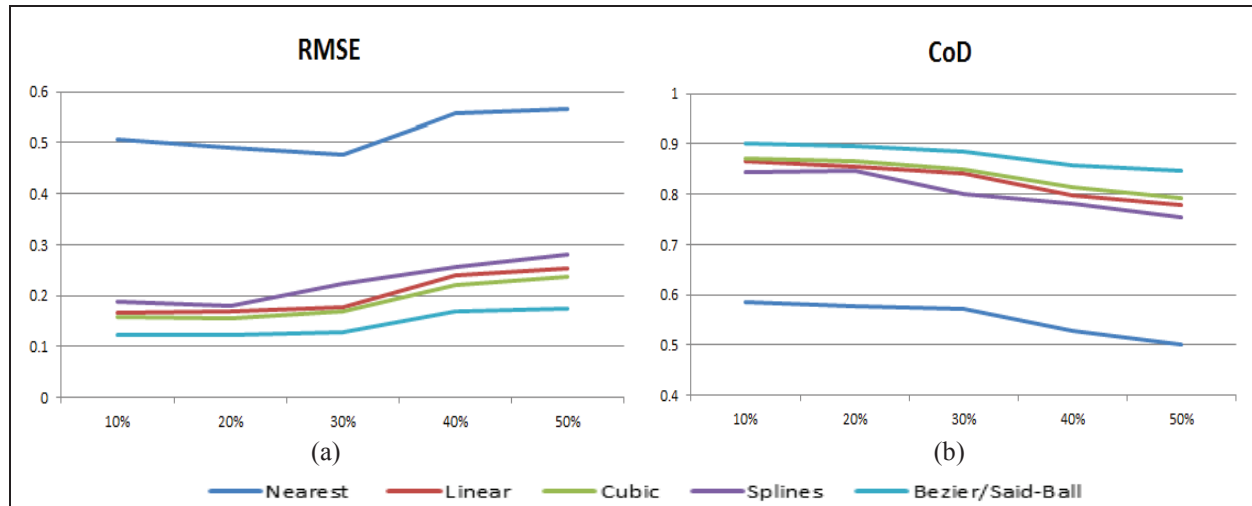
The performance of the piecewise imputation techniques that have been applied to this investigation has been evaluated by the following statistical indicators, Root Mean Square Error (RMSE) and Coefficient of Determination (CoD). Both statistical indicators have exposed the relation between the observed and imputed data as the criteria for a prerequisite evaluation process. Into determining the methods used, which can perform very well into estimating missing values in solar radiation data with respect to the missing value percentage the evaluation process has been made. Five (5) level percentage of simulation missing values data set have been used to examine the goodness-of-fit for each imputation method.

**TABLE 1.** The goodness-of-fit results on each piecewise imputation techniques based on difference amount of missing values percentage

Amount of missing values (%)	Performance Indicators	Imputation Techniques				
		Nearest	Linear	Cubic	Splines	Bezier/Said-Ball
10%	RMSE	0.5070	0.1650	0.1572	0.1888	0.1216
	CoD	0.5841	0.8646	0.8711	0.8451	0.9002
20%	RMSE	0.4897	0.1679	0.1566	0.1791	0.1224
	CoD	0.5774	0.8550	0.8649	0.8454	0.8944
30%	RMSE	0.4774	0.1780	0.1695	0.2237	0.1279
	CoD	0.5720	0.8404	0.8480	0.7994	0.8853
40%	RMSE	0.5577	0.2402	0.2198	0.2573	0.1691
	CoD	0.5282	0.7968	0.8140	0.7823	0.8569
50%	RMSE	0.5673	0.2524	0.2359	0.2813	0.1750
	CoD	0.5018	0.7784	0.7928	0.7529	0.8462



In overall analysis, Bezier interpolation and Said-Ball interpolation well-performed compare to the remains techniques. The remains imputation techniques also present as a good missing value predictor except for nearest interpolation. These environmental can be figured up clearly based on the Figure 3. Here, the Bezier and Said-Ball interpolation does not only provide the similar properties on the mathematical equation but also sharing the output performance. Both approaches have a unique circumstance due to release to share the same performance effect but used two difference intermediate points on the same times. This current situation also happens on the cited research works which proposed by [1]. On the cited research has reported the Bezier and Said-ball interpolation methods had the best imputation method which contributes the same result on the present woks study. At initial observation, Bezier and Said-Ball interpolations are capable and highly potential to become missing value estimator tools. However, the feature's investigation towards the both methods ought to be continues but using difference atmospheric condition due to the advantage that offered into dealing missing values in hourly solar radiation data.



**FIGURE 3.** The statistical analysis comparison on the measured goodness-of-fit between RSME and CoD

On Figure 3, the both statistical indicators have been evaluated the potential on each imputation technique that has employed to predict missing values in hourly solar radiation data. The proposed methods, Bezier and Said-Ball interpolation that used in recent research is based on the cubic approach. Figure 3 has appraised the cubic based methods have been concurred as the better imputation performance. The cubic Bezier and cubic Said-Ball imputation perform best approaches then cubic interpolation is followed. After cubic interpolation, linear interpolation performs better than Spline interpolation. However, the spline interpolation technique also presented as well due to the performance values is slightly below then linear interpolation. Regarding to the Figure 3, the method applied mostly has been demonstrated the missing values prediction performance decreased when the amount of missing values increased. Amount of missing values is a major reason that affects the imputation performance [24].

The lowest imputation method is endower to the nearest interpolation which is the simple technique used to impute missing values. Nearest interpolation only considers the values at the neighbor values, but the remains techniques need consideration on dependent variables and independent variables that applied to generate impute values and then replace the impute values on the absent values which existed in the data set. Thus, nearest interpolation is a technique that not corrects to be used for this research application but the applicant polynomial interpolation methods to impute missing values might be considerable to use for the future works. Based on this research, polynomial interpolation has been obtained as the vital approaches due to the advantage over its potential into imputing missing values in hourly solar radiation data.

## CONCLUSIONS

As conclusion, imputation using polynomial approaches is a reasonable work into estimating missing values in hourly solar radiation data especially for cubic Bezier and Said-Ball interpolation, which have been performed the best methods compare to other proposed imputation methods. The best methods that have been identified in recent

research have constructed by prescribing first-order derivatives and together imply the univariate solar radiation intensity data on the initial and ends of hourly interval data points and other two points at the intermediate interval data points can be predicted. Thus, the piecewise polynomial interpolation is a successful methodology into dealing imputation missing values in hourly solar radiation data.

## ACKNOWLEDGMENTS

The researches would like to thank the Malaysian Meteorological Department due to providing the data for this research. This research has been supported by Ministry of Malaysian Education in term of financial through Exploratory Research Gran Scheme (ERGS) Specification Code Output (S/O) Numbers 12824. Authors on this paper greatly express a very thankful to University Utara Malaysia (UUM) for their favors and effort to complete this work as best as possible.

## REFERENCES

1. A. Saaban, L. Zainudin, and M. N. A. Bakar, "On piecewise interpolation techniques for estimating solar radiation missing values in Kedah," in *Proceedings of 3rd International Conference on Quantitative Sciences and its Applications (ICOQSA 2014)*, AIP Conference Proceedings 1635, edited by H. Ibrahim *et al.* (American Institute of Physics, Melville, NY, 2002), pp. 217–221.
2. I. Moradi, [Energy](#) **34**, 1-6 (2009).
3. C. C. Turrado, M. D. C. M. Lopez, F. S. Lasheras, B. A. R. Gomez, J. L. C. Rolle, and F. J. D. C. Juez, *Sensor* **2013**, **14**, 20382-20399 (2014).
4. S. Ibrahim, I. Daut, Y. M. Irwan, M. Irwanto, N. Gomesh, and Z. Farhana, [Energy Procedia](#) **18**, 1402-1412 (2012).
5. I. Daut, F. Zainuddin, Y. M. Irwan, and A. R. N. Razlina, [Energy Procedia](#) **18**, 1421–1427 (2012).
6. J. Wu and C.K. Chan, [Sol. Energy](#) **85**, 808-817 (2011).
7. C. Simolo, M. Brunetti, M. Maugeri, and T. Nanni, *Int. J. Climat.* **30(10)**, 1564–1576 (2010).
8. D. F. Al-Riza, S. I. Gilani, and M. S. Aris, [Int. J. Environ. Sci and Dev.](#) **2(3)**, 188-193 (2011).
9. N. Khaliliaqdam and A. Soltani, *Int. J. Agric. Crop Sci.* (**IJACS 2012**), 971-978 (2012).
10. J. Fu, D. Jiang, Y. Huang, D. Zhuang, and Y. Wang, *Energies* **2013**, **6**, 2804-2818 (2013).
11. B. G. Akinoglu, [Renewable Energy](#) **1**, 479-497 (1991).
12. G. Y. Shi, T. Hayasaka, A. Ohmura, Z. H. Chen, B. Wang, J. Q. Zhao, H. Z. Che, and L. Xi, [J. Appl. Meteorol. Climatol.](#) **47**, 1006-1016 (2008).
13. F. Besharat, A. A. Deghan, and A. R. Faghih, [Renewable Sustainable Energy Rev.](#) **21**, 798-821 (2013).
14. H. Bulut and O. Buyukalaca, [Appl. Energy](#) **84**, 477–491 (2007).
15. J. Sheffer, *Res. Lett. Inf. Math. Sci.* **3**, 153-160 (2002).
16. A. M. Muzathik, M. Z. Ibrahim, K. B. Samo, and W.B.W. Nik, [Energy](#) **36**, 812-818 (2011).
17. C. Furlan, A. P. de Oliveria, J. Soares, G. Codato, J. F. Escobedo, [Appl. Energy](#) **92**, 240-254 (2012).
18. A. Cagler, C. Yamali, D. K. Baker, and B. Kaftanoglu, *J. Therm. Sci. Technol.*, **33(2)**, 153-142 (2013).
19. I. Korachagaon and V.N. Bapat, [Renewable Energy](#) **41**, 394-400 (2012).
20. A. M. Kalteh and P. Hjorth, [Hydrol. Res.](#) **40.4**, 420-432 (2009).
21. W. J. Tang, K. Yang, J. Qin, C. C. K. Cheng, and J. He, [Atmos. Chem. Phys.](#) **11**, 393-406 (2011).
22. C. Yozgatligil, S. Aslan, C. Iyigun, and I. Batmaz, [Theor. Appl. Climatol.](#) **112**, 143-167 (2013).
23. S. Safi, A. Zeroual, and M. Hassani, [Renewable Energy](#) **27(4)**, 647–666 (2002).
24. H. Junninen, H. Niska, K. Tuppurainen, J. Ruuskanen, and M. Kolehmainen, [Atmos. Environ.](#) **38**, 2895–2907 (2004).