Journal of Engineering and Applied Sciences 13 (9): 2587-2592, 2018
ISSN: 1816-949X

Rational Cubic Spline Interpolation for Missing Solar Data Imputation

[1]Samsul Ariffin Abdul Karim, [2]Mohd Tahir Ismail, [1]Mahmod Othman, [3]Mohd Faris Abdullah, [4]Mohammad Khatim Hasan and [5]Jumat Sulaiman
[1]Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Malaysia
[2]School of Mathematical Sciences, Universiti Sains Malaysia (USM), 11800 Minden, Penang, Malaysia
[3]Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak Darul Ridzuan, Malaysia
[4]Jabatan Komputeran Industri, Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor, Malaysia
[5]Program Matematik dengan Ekonomi, Universiti Malaysia Sabah, Beg Berkunci 2073, 88999 Kota Kinabalu, Sabah, Malaysia

Abstract: Missing data imputation is an important task in statistical and sciences discipline. Solar radiation data obtained from the solar tracker does not complete and some data are missing due to human error in handling the instrument or the failure of the instrument. Thus, missing data imputation can be used to predict and estimate the unknown value of the solar radiation at certain time. This study will estimate the solar radiation by using rational cubic Ball spline function with three parameters. The interpolating rational Ball spline is able to give good result based on quadratic regression model.

Key words: Missing data, imputation, rational cubic Ball, solar radiation, regression model, quadratic

INTRODUCTION

Missing data are common in the scientific problem such as in medicine, electrical, biology, climatology, etc., There are three types of missing data, i.e., data are Missing Completely At Random (MCAR); data are Missing At Random (MAR) and data are missing not a random (MNAR) (Schmitt *et al*., 2015). One of the common strategy in missing data framework is predicting the missing value integrate with statistical analysis. Some examples of imputation methods are fuzzy K-means, K-nearest Neighbor (KNN), etc.

Turrado *et al*. (2014) discussed new imputation algorithm for electrical data loggers. They proposed Multivariate Adaptive Regression Splines (MARS) and compares the performance with Multivariate Imputation by Chained Equations (MICE). From the results, MARS outperform the MICE algorithm. Saaban *et al*. (2016) discussed the estimation of solar radiation missing in Penang, Malaysia. They use piecewise cubic bezier interpolation and develop the sufficient condition for positivity preserving. Radi *et al*. (2015) also discussed the missing data using spatial interpolation. The rainfall data is used for the purpose. They concluded that the Normal Ratio (NR) and Multiple Imputation (MI) give the best estimation for the tested data set. Turrado *et al*. (2015) discussed the missing data imputation of solar radiation data under different atmospheric conditions. They compared the performance of MICE, Inverse Distance Weighting (IDW) and Multiple Linear Regression (MLR). From the numerical results, it was found that MICE give the better result. Norazian *et al*. (2015), utilized linear interpolation and mean imputation techniques to estimate missing value of pollution data PM10 in Seberang Perai, Penang, Malaysia. Karim and Ariffin (2014) discussed the polynomial fitting for global solar radiation. The quadratic polynomial give the better fitting model. Karim (2016a, b) proposed new rational cubic spline with three parameters for data interpolation. In this study, these rational cubic spline will be used to estimate the missing solar radiation data. Three parameters in the description of the rational cubic spline provide greater flexibility in controlling the final estimate value. Some scientific contribution is summarized as follows:

- Missing data imputation using rational cubic spline give higher accuracy and competent with other regression method and cubic spline interpolation

Corresponding Author: Samsul Ariffin Abdul Karim, Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Malaysia

- Three parameters in the description of the rational spline provide flexibility in controlling the final interpolating curve

MATERIALS AND METHODS

This study introduces the new rational cubic spline with two parameters α_i, β_i, i = 1, 2, ..., n-1. The shape control of the rational cubic interpolant also will be discussed in details with numerical examples.

Rational cubic spline interpolant: This study discuss the rational cubic interpolant with three parameters initiated by Karim (2016a, b). For data set $\{(x_i, f_i), i = 0, 1, ..., n\}$ where $x_0<x_1<, ..., <x_n$. Letting $h_i = x_{i+1}-x_i$ $\Delta_i = (f_{i+1}-f_i)/h_i$ and$\theta = (x-x)/h_i$ satisfy $\theta\epsilon[0, 1]$. For each segment on $[x_i, x_{i+1}]$, i = 0, 1, ..., n-1, the rational cubic Ball spline with parameters α_i, β_i and γ_i can be defined as:

$$S(x) = S_i(\theta) = \frac{P_i(\theta)}{Q_i(\theta)} \quad (1)$$

With:

$$P_i(\theta) = \alpha_i f_i (1-\theta)^3 + A_i (1-\theta)^2 \theta + B_i (1-\theta)\theta^2 + F_{i+1}\theta^3$$

And:

$$Q_i(\theta) = \alpha_i (1-\theta)^2 + \beta_i (1-\theta)^2 \theta + \gamma_i (1-\theta)\theta^2 + \theta^2$$

With:

$$A_i = (\alpha_i + \beta_i) f_i + \alpha_i h_i d_i$$
$$B_i = (\gamma_i + 1) f_{i+1} - h_i d_{i+1}$$

The rational cubic spline defined in Eq. 1 satisfies the following C^1 condition:

$$S(x_i) = f_i \text{ and } S(x_{i+1}) = f_{i+1}$$
$$S'(x_i) = d_i \text{ and } S'(x_{i+1}) = d_{i+1}$$

The rational in Eq. 1 can be reformulated as:

$$S(x) = S_i(\theta) = f_i(1-0) + f_{i+1}\theta + h_i\theta(1-\theta)\frac{C_i}{Q_i(\theta)} \quad (2)$$

With:

$$C_i = \alpha_i (1-\theta)(d_i - \Delta_i) + \theta(\Delta_i - d_{i+1}) + (1-\theta)\theta(\gamma_i - \beta_i)\Delta_i$$

From Karim (2015), on each sub-interval $[x_i, x_{i+1}]$, i = 0, 1, ..., n-1, when $\beta, \gamma_i \to \infty$, rational cubic spline in Eq. 1 reduce to the straight line:

$$S(x) \equiv S_i(\theta) = f_i(1-\theta) + f_{i+1}\theta \quad (3)$$

Data collection: The solar radiation data are collected at Universiti Teknologi PTERONAS (UTP). UTP are located at latitude 4°35′2.76″N and longitude 101°04′58.44″E. In our study, the solar radiation is collected every 30 min starting from 7.30 am until 7.30 pm by using solar tracker equip with few sensors and computer to analyst the solar radiation data. In total there are 27 data set. But since the last data is zero, we excluded it in the statistical error measurement. We test the missing data imputation when the data are missing at 10.00 am and 4.30 pm. Figure 1 shows the solar radiation data. Figure 2 shows the instrument to collect solar radiation data in UTP.

Missing imputation data: Missing data imputation usually involving statistical techniques. We propose the framework for missing data imputation by using rational cubic and it involving several steps as described:

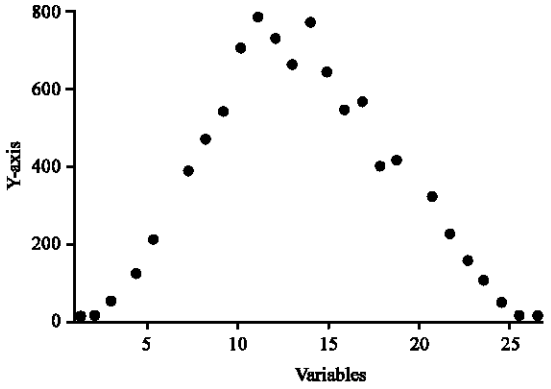

Fig. 1: Solar radiation data with two missing values

Fig. 2: Instrument for solar radiation collected from Karim and Ariffin (2014)

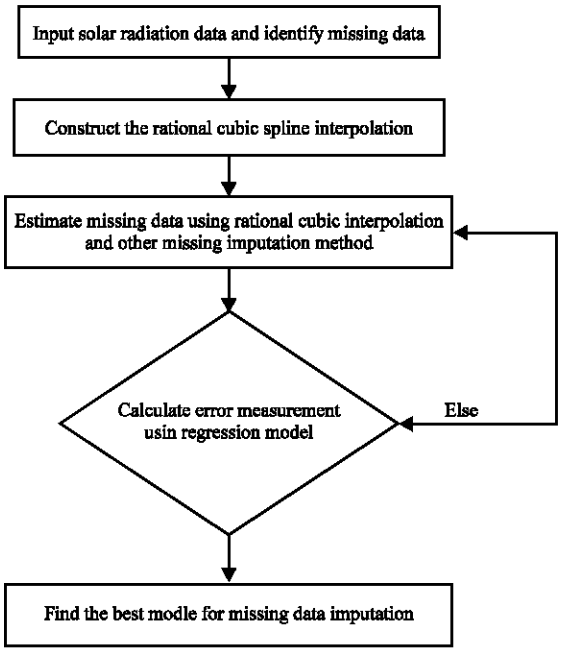

Fig. 3: The framework for missing data imputation

- S_1: interpolate all data by using rational cubic spline given in study
- S_2: for each sub-interval in which the data is missing, we estimate the value by substituting the-value to the rational cubic ball spline
- S_3: repeat S2 with other value of α_i, β_i and γ_i
- S_4: repeat S2 with other existing method such as linear spline, cubic spline and quadratic regression model
- S_5: we calculate the goodness-fit statistics such as RMSE, MAE, MAPE and Theil inequality coefficient based on quadratic regression model
- S_5: compare the performance of all method by comparing the error measurement with respect to the original solar radiation value y. Figure 3 shows the framework for our missing data imputation algorithm

RESULTS AND DISCUSSION

In this study, the missing data imputation from study 2.4 is used to estimate missing solar radiation data. Firstly, we apply rational cubic Ball interpolation with different value of the shape parameters α_i, β_i and γ_i. Then, estimate the missing value at the respective knot (x-value). There are three rational cubic ball, i.e., Rat 1-Rat 3, respectively. We estimate also the missing value by using cubic spline interpolation, linear spline interpolation and regression model. The actual data are obtained from the average of

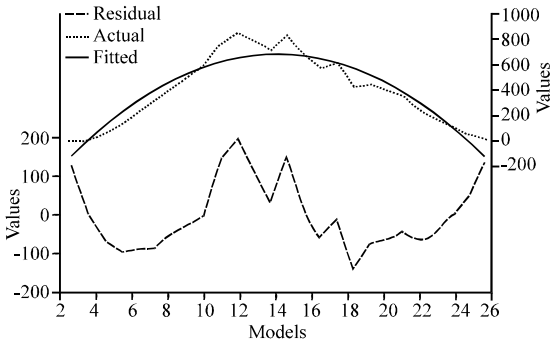

Fig. 4: Model estimation for y; $y = -262.44+140.48x-5.22x^2$

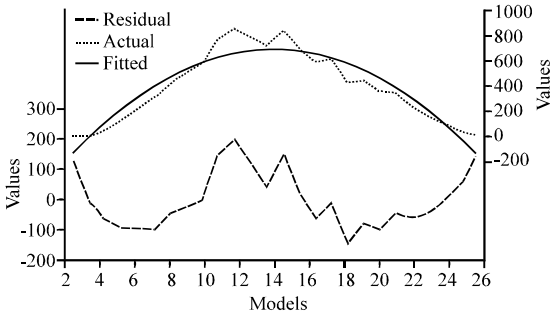

Fig. 5: Model estimation for Cubic Spline (CS); $y = -262.43+140.19x-5.22x^2$

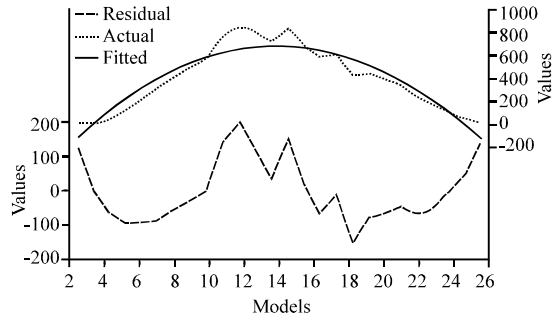

Fig. 6: Model estimation for Linear Spline (LS); $y = -261.37+140.42x-5.22x^2$

the following month. The value at x = 6 and at x = 20 are known as 305.994 and 390.598, respectively. The results are presented in Table 1.

To test the accuracy of the proposed missing imputation method, we use the quadratic regression model as it can see from Fig. 1 that the data show a quadratic pattern. Therefore, the data in Table 1 from column y until regression will be the Dependent Variables (DV) while the first column, x will be the independent variable. The general equation will be given as $DV = c+\alpha x+bx^2$. Figure 4-10 shows the fitted, actual and residual plots for each quadratic regression model. The equations for each model are also given in the figures.

Table 1: Solar radiation data with missing value (highlight in yellow color) with various method

X	Y	Cubic spline	Linear spline	Rat 1	Rat 2	Rat 3	Regression
1	0.074	0.074	0.074	0.074	0.074	0.074	0.074
2	0.450	0.450	0.450	0.450	0.450	0.450	0.450
3	42.738	42.738	42.738	42.738	42.738	42.738	42.738
4	120.258	120.258	120.258	120.258	120.258	120.258	120.258
5	216.606	216.606	216.606	216.606	216.606	216.606	216.606
6	305.994	289.711	315.140	305.832	308.577	304.611	399.630
7	413.674	413.674	413.674	413.674	413.674	413.674	413.674
8	500.144	500.144	500.144	500.144	500.144	500.144	500.144
9	577.940	577.940	577.940	577.940	577.940	577.940	577.940
10	759.376	759.376	759.376	759.376	759.376	759.376	759.376
16	579.800	579.800	579.800	579.800	579.800	579.800	579.800
17	605.784	605.784	605.784	605.784	605.784	605.784	605.784
18	424.178	424.178	424.178	424.178	424.178	424.178	424.178
19	439.746	439.746	439.746	439.746	439.746	439.746	439.746
20	390.598	354.203	388.839	390.528	393.003	389.943	385.210
21	337.932	337.932	337.932	337.932	337.932	337.932	337.932
22	231.506	231.506	231.506	231.506	231.506	231.506	231.506
23	155.516	155.516	155.516	155.516	155.516	155.516	155.516
24	100.070	100.070	100.070	100.070	100.070	100.070	100.070
25	34.468	34.468	34.468	34.468	34.468	34.468	34.468
26	0.644	0.644	0.644	0.644	0.644	0.644	0.644

Table 2: Error measurement based on quadratic regression model

Measurments	Y	Cubic spline	Linear spline	Rat 1	Rat 2	Rat 3	Regression
RMSE	90.900	92.820	90.630	90.910	90.740	90.970	88.090
MAE	74.540	76.230	74.290	74.540	74.390	74.600	69.880
MAPE	7507.540	7540.130	7449.400	7508.390	7496.340	7514.670	7146.720
Theil	0.096	0.098	0.095	0.096	0.095	0.096	0.092

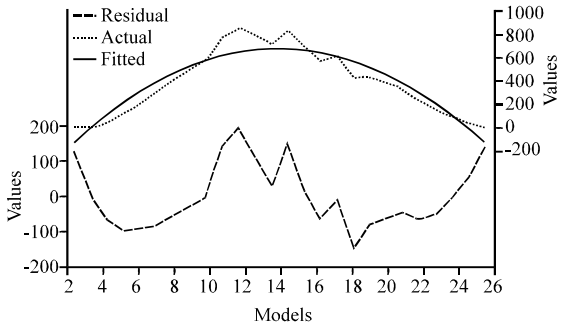

Fig. 7: Model estimation for rat 1; y = -262.45+140.48x -5.22x^2

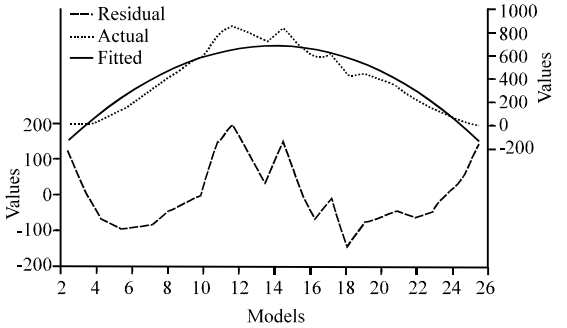

Fig. 9: Model estimation for rat 3; y = -262.56+140.48x -5.22x^2

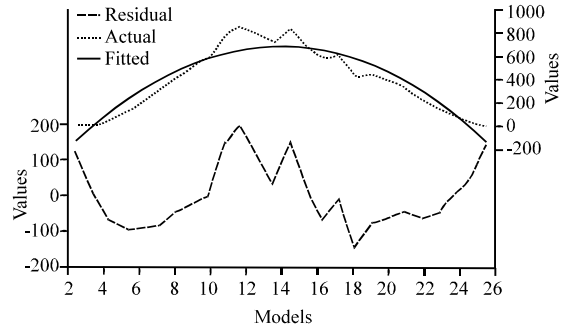

Fig. 8: Model estimation for rat 2; y = -262.28+140.49x-5.22x^2

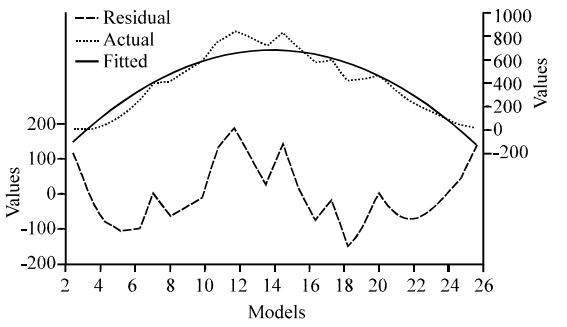

Fig. 10: Model estimation for Regression (Reg.); y = -262.86+140.75x-5.23x^2

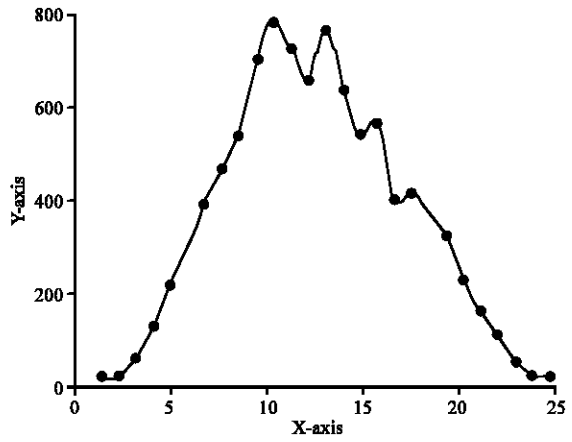

Fig. 11: Interpolating curve using rational cubic interpolation (Rat 1)

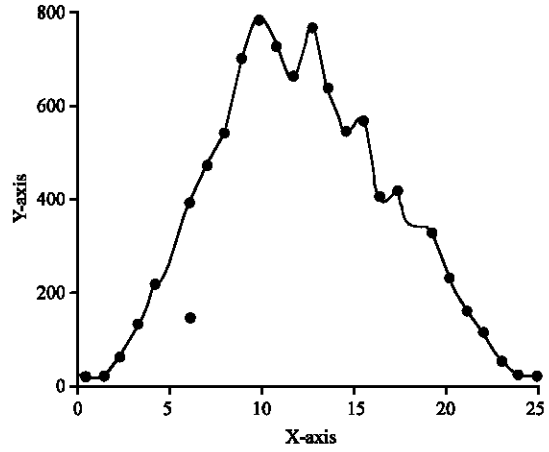

Fig. 12: Interpolating curve using cubic spline interpolation

Based on Fig. 4-10, it appears that the fitted observations from the quadratic regression model manage to follow the same trend as the actual observations. In order to find the best method that produce the missing values nearest to the real values, y the error measurement, i.e., RMSE, MAE, MAPE and Theil coefficient are used. From Table 2, it can be seen clearly that the missing values obtained from Rat 1 give the best result, since, it has the value of RMSE, MAE, MAPE and Theil very close to the actual values of y (real solar radiation data).

Moreover, Fig. 11 shows the interpolating curve for Rat 1 with α_i, $\beta_i = 1$ and $\gamma_i = 0.36$ $\gamma_5 = 0.36$ and $\gamma_{18} = 3.9$. Meanwhile Fig. 12 shows the interpolating curve using cubic hermite spline interpolation.

CONCLUSION

This study discussed the missing data imputation by using rational cubic spline with three parameters. From all numerical results, the unknown value, i.e., the missing solar radiation data has been predicted with higher accuracy as indicated in the less error and very closed to the original value for the data taking from different month. To improve the missing data prediction, the free parameters in the description of the rational cubic spline can be altered according to what user need. The proposed scheme can be used to estimate the missing imputation data for other type of data set.

RECOMMENDATION

Further study will be emphasized more on other missing data imputation techniques and compare the performance against the rational cubic spline interpolation.

ACKNOWLEDGEMENT

This research is fully supported by Universiti Teknologi PETRONAS (UTP) through STIRF grant: 0153AA-D91 including Mathematica software.

REFERENCES

Karim, A. and S. Ariffin, 2014. Global solar radiation modeling using polynomial fitting. Appl. Math. Sci., 8: 367-378.

Karim, S.A.A., 2015. Shape preserving by using rational cubic ball interpolant. Far East J. Math. Sci., 96: 211-230.

Karim, S.A.A., 2016a. Data interpolation using rational cubic Ball spline with three parameters. AIP. Conf. Proc., 1787: 080023-1-080023-8.

Karim, S.A.A., 2016b. Positivity preserving interpolation by using rational cubic ball spline. Jurnal Teknologi, 78: 141-148.

Norazian, M.N., Y.A. Shukri, R.N. Azam and A.M.M. Al Bakri, 2008. Estimation of missing values in air pollution data using single imputation techniques. Sci. Asia, 34: 341-345.

Radi, N.F.A., R. Zakaria and M.A.Z. Azman, 2015. Estimation of missing rainfall data using spatial interpolation and imputation methods. AIP Conf. Proc., 1643: 42-48.

Saaban, A., M.Z. Zainuddin and M.N.A. Bakar, 2016. Piecewise positivity preserving cubic bezier interpolation for estimating solar radiation missing vale in Penang, Malaysia. J. Math. Stat., 12: 302-307.

Schmitt, P., J. Mandel and M. Guedj, 2015. A comparison of six methods for missing data imputation. J. Biomet. Biostat., 6: 1-6.

Turrado, C.C., M.D.C.M. Lopez, F.S. Lasheras, B.A.R. Gomez and J.L.C. Rolle *et al*., 2014. Missing data imputation of solar radiation data under different atmospheric conditions. Sensors, 14: 20382-20399.

Turrado, C.C., S.F. Lasheras, J.L. Calvo-Rolle, A.J. Pinon-Pazos and F.J.D.C. Juez, 2015. A new missing data imputation algorithm applied to electrical data loggers. Sensors, 15: 31069-31082.