Fitting Rainfall Data by Using Cubic Spline Interpolation

Irham Azizan[1] *, Samsul Ariffin* Bin Abdul Karim[2,*] and *S. Suresh* Kumar Raju[3]

[1]Department Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 32610 Seri Iskandar, Perak Darul Ridzuan, Malaysia.
[2]Fundamental and Applied Sciences Department and Centre for Smart Grid Energy Research (CSMER), Institute of Autonomous System, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 32610 Seri Iskandar, Perak DR, Malaysia.
[3]Department of Mathematics and Statistics, King Faisal University, Saudi Arabia.

Abstract. This study discusses the application of two cubic spline i.e. natural and not-a-knot end boundary conditions to visualize and predict the rainfall data. The interpolation and the analysis of the rainfall data will be done on a monthly basis by using the MATLAB software. The rainfall data is obtained from Malaysia Meteorology Department for Ipoh and Petaling Jaya in year 2014 and 2015. The interpolating curves are then being compared and if there is any negative value on the interpolating curve on some sub-interval, that part will be replaced by using the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP). We discuss the missing data imputation by using both splines.

1 Introduction

Since Malaysia is located near the equator, there are always climate change such as hot and rainy season throughout the year. The weather is constantly changing every period of months and different states have different amount of rainfall. These are due to the factors such as wind speed and wind direction, the humidity, the temperature (max, min and average) in Malaysia. Due to this constant change, the amount of the rainfall need to be recorded and collected as they have significant impact on the environment such as the plants life, food and water supplies [1]. The collection of data is done by the Meteorology department in the respective states and these data then can be interpolated so that the amount of the rainfall for the next month can be calculated and predicted. This data is important especially to the farmers and gardeners as they need to monitor their plants during the planting and harvesting seasons [1]. Throughout the year, Malaysia has undergone cold stormy rain for every period of months and this affects the ecosystems and people who depends on them such as the farmers and gardeners. Therefore, a rainfall data monitoring is being done by the Meteorology department in each state so that the rainfall can be observed for a specific period. Therefore, the rainfall data obtained will not be uniform as Malaysia has its rain season for a certain period throughout the year. Different states will have different period due

---

* Corresponding author: samsul_ariffin@utp.edu.my

to the Northeast Monsoon causing heavy rainfall and the Southwest Monsoon that causes hot weather [2]. As for an example, the eastern coast states of Peninsular Malaysia such as Kelantan and western Sarawak will have a heavy rainfall due to this wind season. Due to this constant change of weather, the rainfall data will not be uniform and to solve this, the cubic spline interpolation is used for the numerical analysis. Cubic spline interpolation is chosen as its error margin is smaller and it gives a smoother interpolation compared to the other interpolation methods [3]. There are two boundary conditions in cubic spline interpolation i.e. natural and not-a-knot. Karim et al. [4] had also used linear, Piecewise Cubic Hermite interpolating Polynomial (PCHIP) and cubic spline interpolation to interpolate the petroleum engineering data.

Even though the PCHIP has lower order of continuity compared to the cubic spline, Abbas et al. [5] stated that it is still useful in some case as its interpolation value will always be positive. Work in this study is relevant to the alternative energy i.e. energy harvesting from hydroelectric power that depends on volume of the rainfall. The relevant agency can use the proposed model to predict the amount of the rainfall volume at certain location.

In this study, we will use two types of cubic spline interpolation i.e. natural splines and not-a-knots splines. Both splines will be compared and if there are any negative values on some sub-interval, the interpolating curve on that interval will be replaced by PCHIP – which is a guarantee to produce positive interpolating but at the cost of the interpolating curves being not smooth as well as the curves tend to tight on certain interval. Finally, we consider the missing rainfall data imputation. Overall, not-a-knot cubic spline interpolation is better than natural cubic spline interpolation in terms of the accuracy of the missing data imputation as well as visually pleasing results.

2 Introduction C^2 cubic spline interpolation

There are many variants of cubic spline interpolation [6]. The two most common are cubic spline interpolation with natural and not-a-knot end conditions. Figs. 1 and 2 show the example of natural, not-a-knot spline and PCHIP for data given in Tables 1 and 2 respectively.

Table 1. Temperature as a function of resistance

i	1	2	3	4
R (ohm)	451.1	636	911.3	1101
T (°C)	50.128	40.120	30.131	25.113

Table 2. An example of a set of data

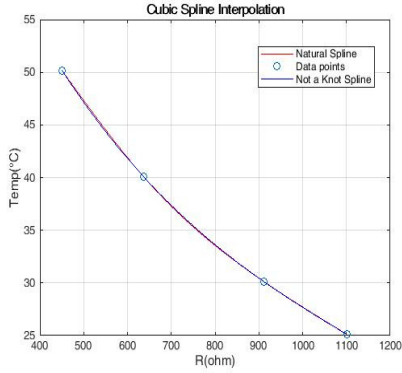| i	1	2	3	4
x	1	2	3	4
y	220.8	42.4	175.2	573.2

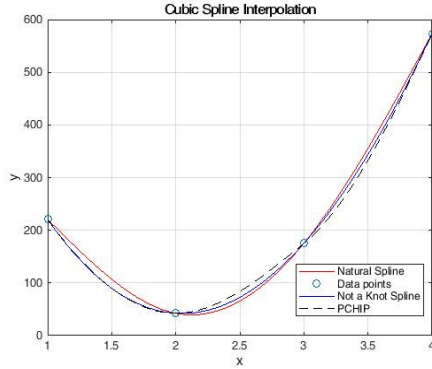**Fig. 1.** Spline interpolation for in Table 1.

Fig. 2. Spline interpolation for in Table 2.

3 Rainfall data prediction

As a measurement to validate the prediction value by using cubic spline interpolation i.e. natural, not-a-knots and PCHIP, the true error relative percentage is used. After we obtain the interpolating curve for all rainfall data at Ipoh and Petaling Jaya for 2014 and 2015, we observed to see if both boundary conditions are suitable or not for the rainfall data in that specific location and time. In addition, a real-time situation is also created to analyze the splines where one month is made to be missing which is May and the missing month rainfall data will be found using the interpolation. Then, the missing month rainfall data is compared to its actual value by calculating the error percentage and the error percentage that has an error less than 20% is accepted. The formula used for the error percentage is given as

$$\text{True relative percentage error} = \left| \frac{\text{Actual value - interpolation value}}{\text{Actual value}} \right| \times 100\% \qquad (1)$$

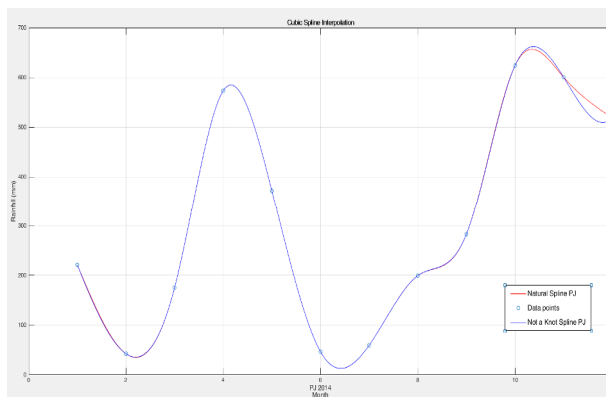Cubic spline interpolation for Petaling Jaya 2014 rainfall data

Fig. 3. Natural and not-a-knot spline of rainfall data for Petaling Jaya 2014

Table 3. Analysis of the mid-month of Petaling Jaya 2014 rainfall data

Month	Rainfall for PJ 2014	
	Natural Spline	Not-a-Knot Spline
1.5	111.4	109.3
2.5	52.7	53.25
3.5	402.5	402.5
4.5	540.2	540.3
5.5	182.2	182.2
6.5	13.97	14.01
7.5	138.1	137.9
8.5	218.4	218.9
9.5	459	457.1
10.5	652	659.1
11.5	553.8	527.5

Table 4. Analysis of one missing month of Petaling Jaya 2014 rainfall data

Month	PJ 2014				
	Rainfall			Error Percentage	
	Actual	Natural Spline	Not-a-Knot Spline	Natural Spline	Not-a-Knot Spline
1	220.8	220.8	220.8	0	0
2	42.4	42.4	42.4	0	0
3	175.2	175.2	175.2	0	0
4	573.2	573.2	573.2	0	0
5	372.2	403	403.1	8.28	8.30
6	46.6	46.6	46.6	0	0
7	59.4	59.4	59.4	0	0
8	199.2	199.2	199.2	0	0
9	282.6	282.6	282.6	0	0
10	624	624	624	0	0
11	600.6	600.6	600.6	0	0
12	518.8	518.8	518.8	0	0

From Fig. 3, it can be seen that both boundary conditions are suitable for the rainfall data interpolation at Petaling Jaya in 2014. Furthermore, both splines gave positive values which are realistic for a rainfall data and based on these interpolation, the missing rainfall data in May is analyzed and based on the error percentage for both splines, we found that both splines are suitable for missing rainfall data imputation. Tables 3 and 4 summarize the prediction.
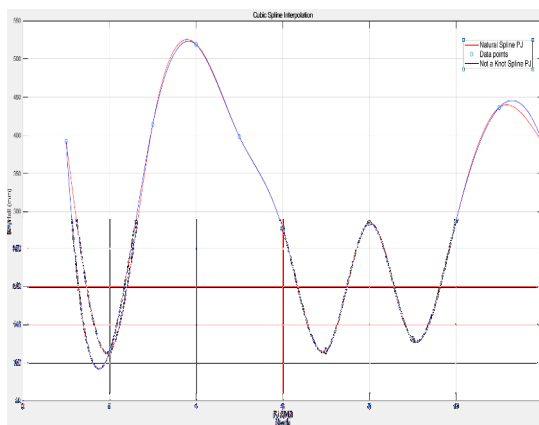
Fig. 4. Natural and not-a-knot spline of rainfall data for Petaling Jaya 2015

Based on Fig. 4, the values of the rainfall data from both splines are positive and they are still applicable for the rainfall data. However, a further analysis needs to be done on a day to day basis for accuracy purposes as the values of both splines has a lot of differences with the actual data. Other than that, the error percentage for the missing month which is May is acceptable as both splines have an error of 14.11% and 14.97% for the natural spline and not-a-knot spline respectively.

Table 5. Analysis of mid- month of Petaling Jaya 2015 rainfall data

Month	Rainfall for PJ 2015	
	Natural Spline	Not-a-Knot Spline
1.5	191.3	121.4
2.5	232.2	250.9
3.5	512.3	507.3
4.5	466	467.3
5.5	347	346.6
6.5	165.4	165.5
7.5	204.3	204.3
8.5	215.9	215.7
9.5	172	172.9
10.5	385.9	382.9
11.5	429.7	441.1

Table 6. Analysis of one missing month of Petaling Jaya 2015 rainfall data

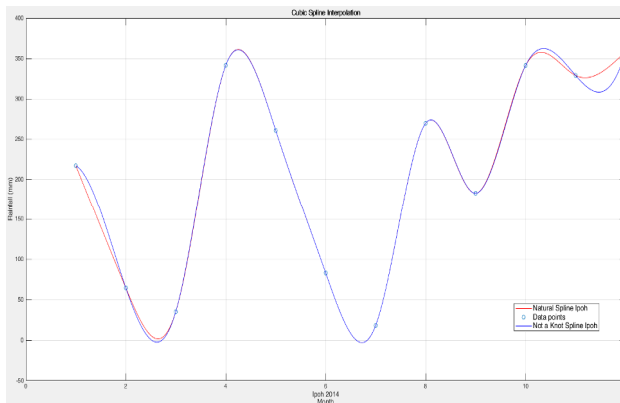|	PJ 2015				
Month	Rainfall			Error Percentage	
	Actual	Natural Spline	Not-a-Knot Spline	Natural Spline	Not-a-Knot Spline
1	392.2	392.2	392.2	0	0
2	113.8	113.8	113.8	0	0
3	414.4	414.4	414.4	0	0
4	519	519	519	0	0
5	398.2	454.4	457.8	14.11	14.97
6	276.6	276.6	276.6	0	0
7	115.6	115.6	115.6	0	0
8	283.6	283.6	283.6	0	0
9	130.2	130.2	130.2	0	0
10	287.2	287.2	287.2	0	0
11	436.4	436.4	436.4	0	0
12	390.2	390.2	390.2	0	0

Fig. 5. Natural and not-a-knot spline of rainfall data for Ipoh 2014
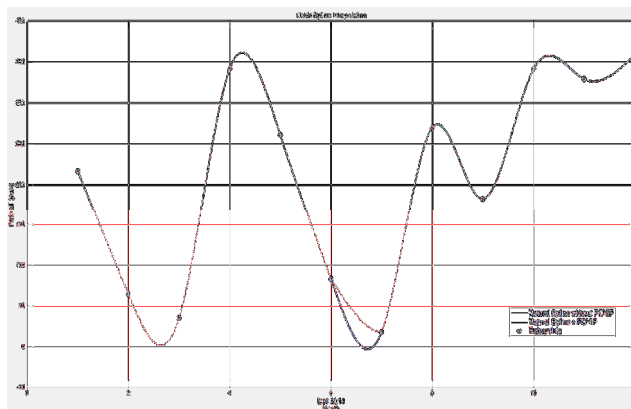

Fig. 6. PCHIP added to compensate the negative value of the natural and not-a-knot spline

Table 7. Analysis of mid-month of Ipoh 2014 rainfall data

Month	Rainfall for Ipoh 2014	
	Natural Spline	Not-a-Knot Spline
1.5	140.1	159.1
2.5	5.935	0.8351
3.5	192.9	194.3
4.5	345.8	345.5
5.5	169.8	169.9
6.5	8.335	8.342
7.5	154.1	154
8.5	235.9	236.3
9.5	248	246.3
10.5	353.5	359.9
11.5	332.4	308.5

From Fig. 5, even though almost all the rainfall data from both splines are fine, there are still some part of the splines that need to be adjusted. These can be seen on the interval between February and March for the natural spline and the interval between June and July for both splines. As can be seen from the graph, the values for these intervals are in a negative value and these are impossible as the rainfall data is always in positive. In the case for the February and March interval, the not-a-knot spline should be chosen for the numerical analysis as its value is still in the positive values. Meanwhile, the adjustment that is made for the June and July interval is that different interpolation method, PCHIP is used as seen in Figure 6. Table 5 until 7 summarize the prediction.
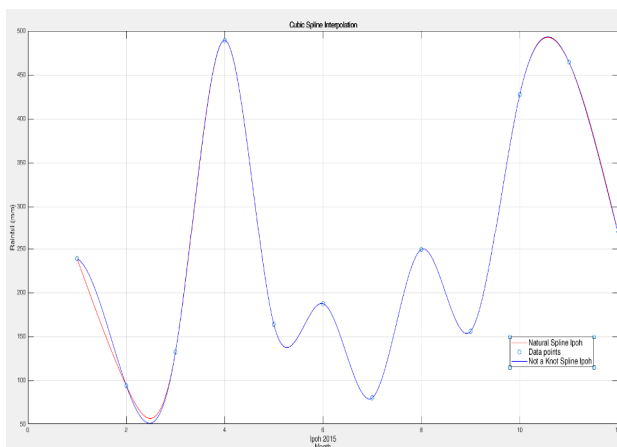

Fig. 7. Natural and not-a-knot spline of rainfall data for Ipoh 2015

Table 8. Analysis of one missing month of Ipoh 2014 rainfall data

Ipoh 2014					
Month	Rainfall			Error Percentage	
	Actual	Natural Spline	Not-a-Knot Spline	Natural Spline	Not-a-Knot Spline
1	217	217	217	0	0
2	64.6	64.6	64.6	0	0
3	35.2	35.2	35.2	0	0
4	341.4	341.4	341.4	0	0
5	260.8	322.2	321.3	23.54	23.20
6	83	83	83	0	0
7	18.2	18.2	18.2	0	0
8	269.6	269.6	269.6	0	0
9	182.8	182.8	182.8	0	0
10	341.2	341.2	341.2	0	0
11	329	329	329	0	0
12	358.4	358.4	358.4	0	0

In Figure 7, the interpolation for the rainfall data in Ipoh 2015 for both splines has no problem since both values are positive and it is still within the acceptable range for a rainfall data type. However, the analysis on a missing month i.e. May has shown that the error percentage for the missing month is very high and this is not a good value as error percentage must be at least below 20%. In order to solve this problem, more sample points could be added to decrease the error percentage and improve the accuracy of the prediction. The results show in this study can be further extended to the development of the high accuracy of the prediction-based model. This will give the relevant agency the adequate prediction model in order to forecast the rainfall at certain location. Tables 8 and 9 summarize the prediction as well as the missing data imputation.

Table 9. Analysis of one mid-month of Ipoh 2015 rainfall data

Month	Rainfall for Ipoh 2015	
	Natural Spline	Not-a-Knot Spline
1.5	162	182.2
2.5	56.53	51.14
3.5	352	353.4
4.5	353.8	353.4
5.5	150.2	150.3
6.5	125.9	125.8
7.5	166.2	166.2
8.5	198	198.1
9.5	268.9	268.7
10.5	491.7	492.3
11.5	380.8	378.5

Discussions and conclusion

In this study, three types of cubic splines are investigated in detail for rainfall data interpolation. They are natural spline, not-a-knot spline and PCHIP. We test the natural and not-a-knot spline to interpolate the rainfall data at two locations: Petaling Jaya and Ipoh for year 2014 and 2015. Both splines are capable to interpolate as well as give the results. However, from Figure 5, for rainfall data at Ipoh, 2014, both splines give negative value at the sub-interval of June and July. Since the rainfall data is positive (any negativity is meaningless), therefore PCHIP is embedded into the interval. This will produce positive interpolating curve. A new prediction model for rainfall data is currently investigated by the authors. Early results indicate some interesting results. We intend to complete the study and report the main outcome soon.

References

1. A. Reynolds, "Why Is a Rain Gauge Important?," 25 April 2017. [Online]. Available: https://sciencing.com/rain-gauge-important-6611576.html.
2. M. Cavendish, World and Its Peoples: Eastern and Southern Asia (2008).
3. T. Young and M. J. Mohlenkamp, Introduction to Numerical Methods and MATLAB Programming for Engineers, Colombus: Ohio University (2017).
4. S. Abdul Karim, M. Mohd Rosli and M. Mohd Mustafa, Applied Mathematical Sciences, vol. 8, pp. 5083-5098 (2014).
5. M. Abbas, A. A. Majid and J. M. Ali. ScienceAsia, 39, pp. 208-213 (2013).
6. K. Wang, InSight: Rivier Academic Journal, 9, pp. 2 (2013).