Interpolation on Lincoln Weather

Khang Phan

December 2018

Contents

List of Figures

1   Introduction

2   Data

The data is collected from National Climatic Data Center. It contains information in several data station across Nebraska from January 2010 to 2018. For the investigation, we use the data in 2011 at station USW00014939, which has 366 data points.

3   Interpolation Method

In this investigation, we measure four different interpolation method: Linear Spline, Quadratic Spline, Cubic Spline, and Best Fit Polynomial.

3.1   Linear Spline

Each piece of the linear splines can be found using a simple equation:

$$S_i(x) = y_i + \frac{y_{i+1} - y_i}{t_{i+1} - t_i}(x - t_i), \qquad i = 0, 1, \cdots, n - 1.$$

Figure 1: Linear Spline

3.2   Quadratic Spline

$$z_0 = 0$$

$$z_{i+1} = -z_i + 2\left(\frac{y_{i+1} - y_i}{t_{i+1} - t_i}\right)$$

$$Q_i(x) = \frac{z_{i+1} - z_i}{2(t_{i+1} - t_i)}(x - t_i)^2 + z_i(x - t_i) + y_i$$

Figure 2: Quadratic Spline

3.3 Cubic Spline

INPUT n; $x_0, x_1, \ldots, x_n$; $a_0 = f(x_0), a_1 = f(x_1), \ldots, a_n = f(x_n)$.

OUTPUT a_j, b_j, c_j, d_j for $j = 0, 1, \ldots, n-1$.
(Note: $S(x) = S_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3$ for $x_j \le x \le x_{j+1}$.)

Step 1 For $i = 0, 1, \ldots, n-1$ set $h_i = x_{i+1} - x_i$.

Step 2 For $i = 1, 2, \ldots, n-1$ set

$$\alpha_i = \frac{3}{h_i}(a_{i+1} - a_i) - \frac{3}{h_{i-1}}(a_i - a_{i-1}).$$

Step 3 Set $l_0 = 1$;   *(Steps 3, 4, and 5 and part of Step 6 solve a tridiagonal linear system using a method described in Algorithm 6.7.)*

$\mu_0 = 0$;
$z_0 = 0$.

Step 4 For $i = 1, 2, \ldots, n-1$
set $l_i = 2(x_{i+1} - x_{i-1}) - h_{i-1}\mu_{i-1}$;
$\mu_i = h_i / l_i$;
$z_i = (\alpha_i - h_{i-1}z_{i-1})/l_i$.

Step 5 Set $l_n = 1$;
$z_n = 0$;
$c_n = 0$.

Step 6 For $j = n-1, n-2, \ldots, 0$
set $c_j = z_j - \mu_j c_{j+1}$;
$b_j = (a_{j+1} - a_j)/h_j - h_j(c_{j+1} + 2c_j)/3$;
$d_j = (c_{j+1} - c_j)/(3h_j)$.

Step 7 OUTPUT $(a_j, b_j, c_j, d_j$ for $j = 0, 1, \ldots, n-1)$;
STOP.

Figure 3: Cublic Spline

3.4 Best Fit Polynomial

Polynomial function:

$$f(x) = \sum_{i=0}^{k} a_i x^i = a_0 + \ldots + a_k * x^k$$

Optimize function:

$$error = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

4 Experiment

4.1 Sampling method

We ramdomly select sets of points to use as input for our interpolation methods. Other points are used to measure the accuracy. The sets always contain the data on January 1, 2011 and January 1, 2012, so all

missing points are in the range. The size of the sets include 330, 300, 270, 240, 73, and 37.

4.2   Graph

Following are the graph of different methods used on a set of input sized 270. The green line indicates the actual data, and the red line is my predicted function.

The first three graphs are Linear Spline, Quadratic Spline, and Cubic Spline respectively. From the graph, we can see clearly Quadratic Spline does not work in this case. It goes far above the below the actual data points. On the other hand, Cubic Spline and Linear Spline are much more accurate. They are much closer to the actual data. Also, they are visually similar to each other.
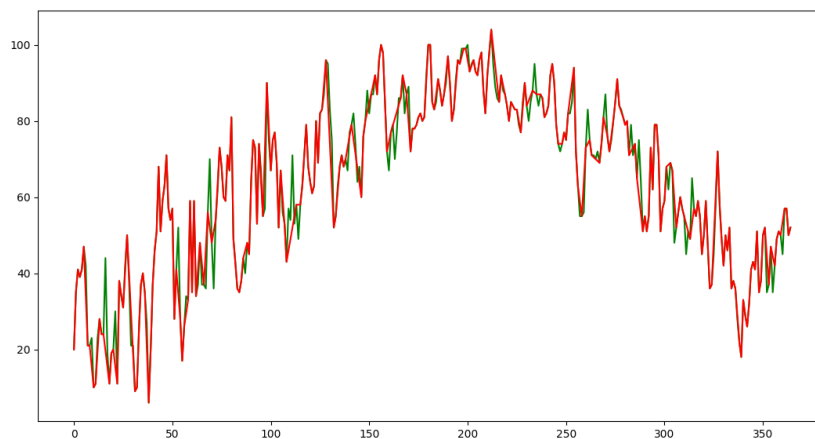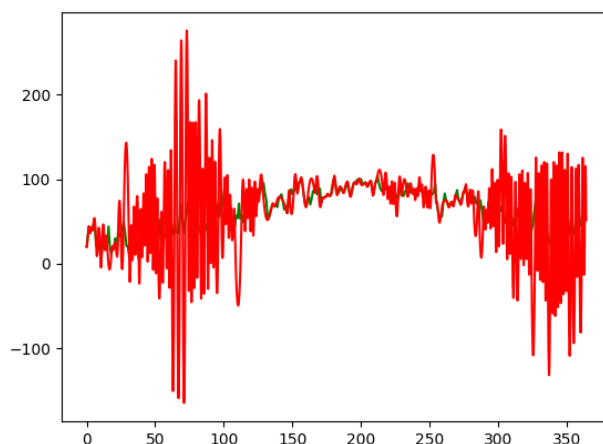

Figure 4: Linear Spline
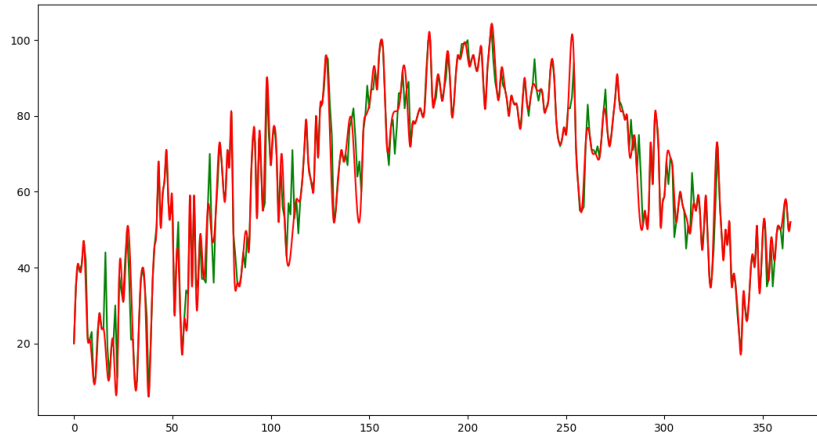

Figure 5: Quadratic Spline

Figure 6: Cubic Spline

Next figures are graphs of degree one to nine polynomials having smallest sum of square errors. The figures give us the trend of temperature. From the curves, we have some idea when the temperature should go up or down. From the graphs, we can see the difference between linear function and the rest. When comparing polynomial degree two to nine, we could not see big changes between any two consecutive functions.
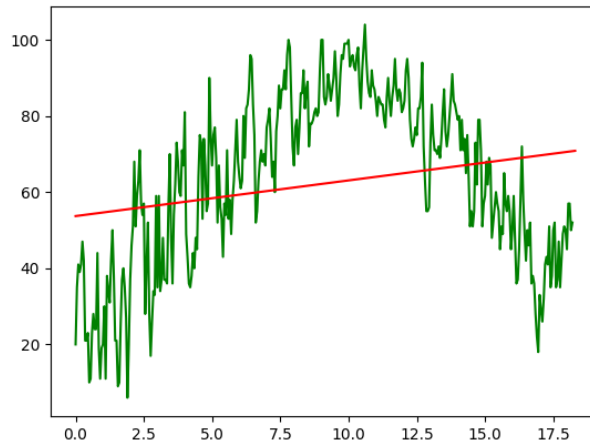

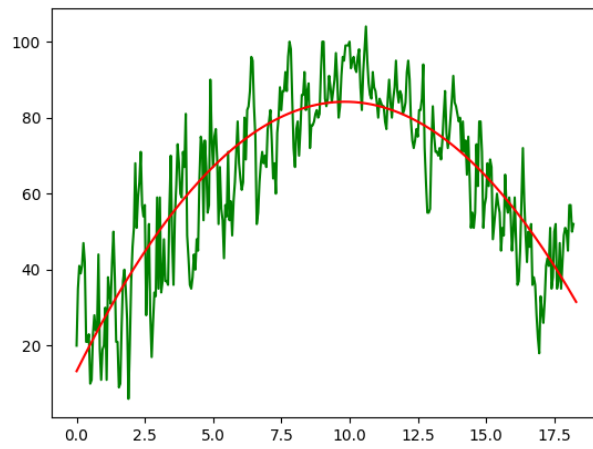
Figure 7: Degree one polynomial
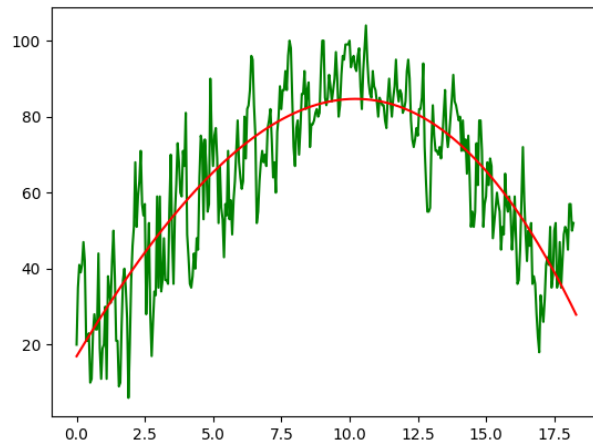
Figure 8: Degree two polynomial

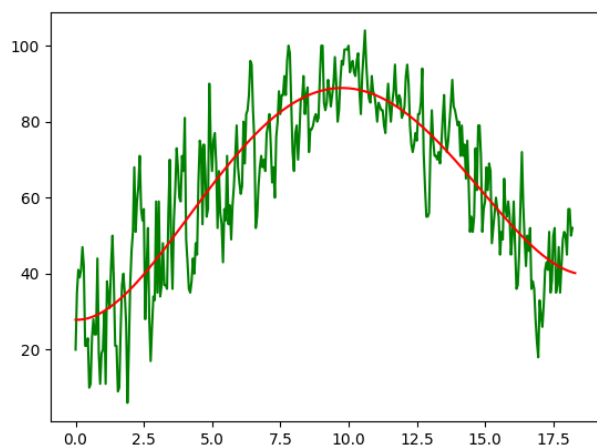Figure 9: Degree three polynomial
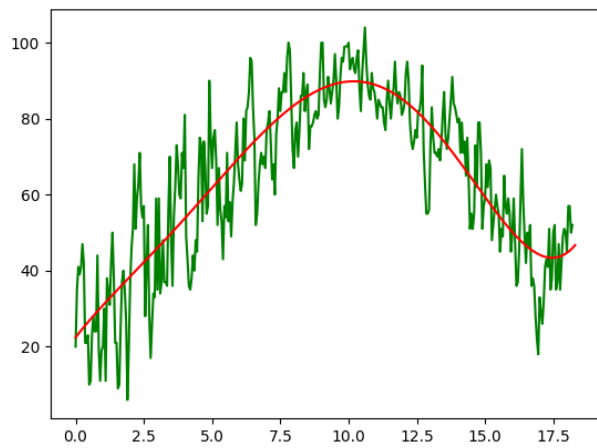
7

Figure 10: Degree four polynomial

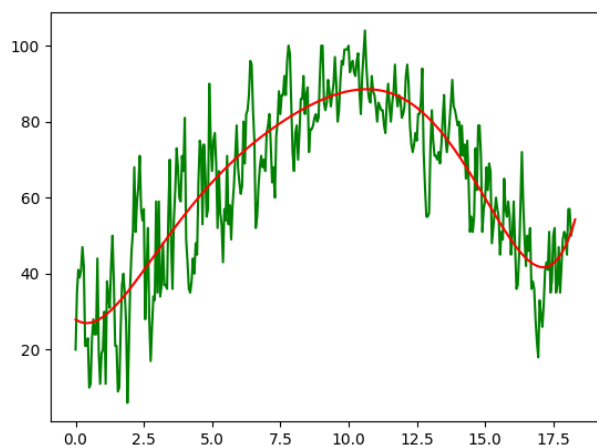Figure 11: Degree five polynomial

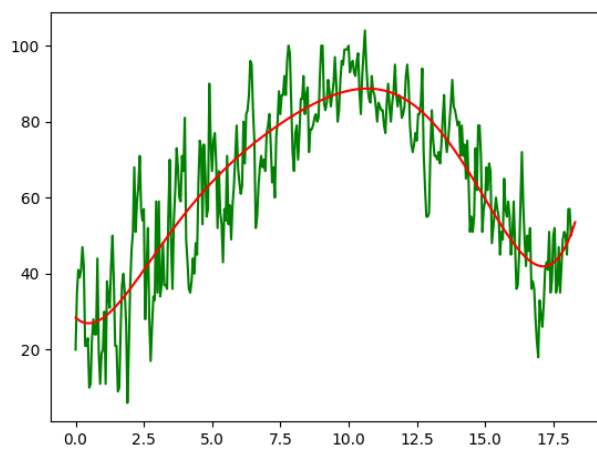Figure 12: Degree six polynomial
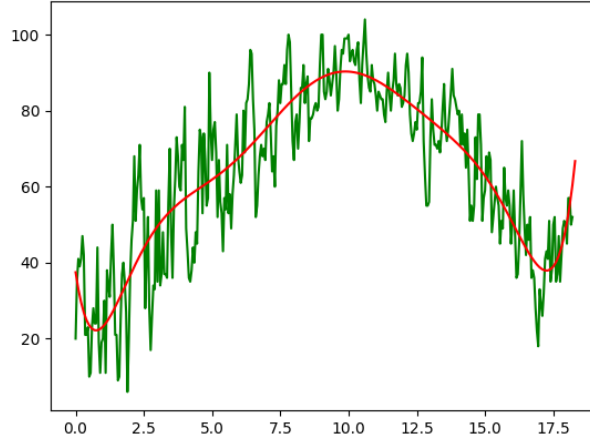

Figure 13: Degree seven polynomial
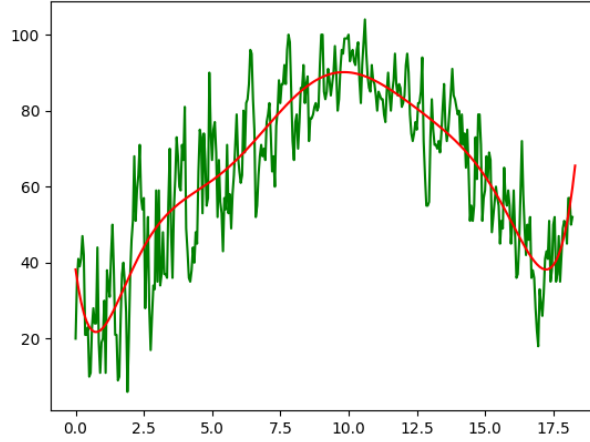
Figure 14: Degree eight polynomial

Figure 15: Degree nine polynomial

4.3 Aggregation

To compare the accuracy of above methods, we randomly sample data points and run interpolations on them. Then we use the mean of square errors to determine their accuracy. Figure 16 shows the results after 100 runs of each size. We can see, for each method (i.e. each column), as the sample size decreases, the errors increases.

Comparing the piecewise interpolation, Quadratic Spline error is significantly larger then Linear or Cubic. Linear Spline and Cubic Spline is very similar, as we can see from above graph; as Linear seems to be a little better. Comparing polynomials, linear has much higher errors, while other degree from two to nine are close. As degree gets higher, the error rates decrease. We also notice that as the sample size decreases, error terms of Piecewise interpolation method increases much faster than Polynomial functions.

Sample	Spline			Polynomial								
size	Linear	Quadratic	Cubic	Linear	2	3	4	5	6	7	8	9
330	2.33	39.48	2.69	21.62	11.29	10.93	10.55	10.51	10.51	9.81	9.81	9.67
300	3.24	52.97	3.77	21.63	11.29	10.94	10.56	10.52	10.52	9.82	9.83	9.69
270	4.02	66.16	4.69	21.64	11.30	10.94	10.57	10.53	10.54	9.84	9.84	9.70
240	4.79	67.00	5.71	21.65	11.31	10.96	10.59	10.55	10.56	9.86	9.87	9.73
73	9.41	130.25	14.64	21.86	11.49	11.17	10.82	10.82	10.87	10.29	10.34	10.20
37	10.88	131.92	18.94	22.22	11.70	11.40	11.12	11.22	11.42	10.97	11.47	11.50

Figure 16: Result

5 Conclusion

After the experiments, our first impression is that Linear Spline seems to be the best method in general. Cubic Spline would also work very well, but it requires more complex implementation. When the sample size is small, we can use Polynomial functions, as they give us information about the temperature trend with accuracy comparable of Linear/Cubic Spline give.

6 Reference