APPLICATION OF NON PARAMETRIC BASIS SPLINE (B-SPLINE) IN TEMPERATURE FORECASTING

[1]Rezzy Eko Caraka, [2]Alvita Rachma Devi
[1,2]Department of Statistics, Diponegoro University, Semarang
Email: rezzyekocaraka@gmail.com

ABSTRACT

"Weather is important but hard to predict"—lay people and scientists alike will agree. The complexity of system limits the knowledge about it and therefore its predictability even over a few days. It is complex because many variables within the Earth's atmosphere, such as temperature and they do so nonlinearly. B-spline as a basis for one-dimensional regression and we extend this paper by using B-spline to construct a basis for bivariate regression. This construction gives a basis in two dimensions with local support and hence a fully flexible family of fitted mortality surfaces one of the principal motivations behind the use of B-spline as the basis of regression is that it does not suffer from the lack of stability that can so bedevil ordinary polynomial regression. The essential difference is that B-spline have local non-zero support in contrast to the polynomial basis for standard regression. The optimal B-Spline models rely on the optimal knots that has a minimum Generalized Cross Validation (GCV)

Keywords: Temperature, B-Spline, Generalized Cross Validation, non-parametric

INTRODUCTION

Far from being future fear, global warming is happening now, and scientists have evidence that humans are to blame. For decades, cars and factories have spewed billions of tons of greenhouse gases into the atmosphere, and these gases caused temperatures to rise between 0.6°C and 0.9°C (1.08°F to 1.62°F) over the past century. The rate of warming in the last 50 years was double the rate observed over the last 100 years. Temperatures are certain to go up further. Several studies have looked at the potential impact of climate change.

Riebeek explained Global warmth begins with sunlight. When light from the Sun reaches the Earth, roughly 30 percent of it is reflected back into space by clouds, atmospheric particles, reflective ground surfaces, and even ocean surf. The remaining 70 percent of the light is absorbed by the land, air, and oceans, heating our planet's surface and atmosphere and making life on Earth possible. Solar energy does not stay bound up in Earth's environment forever. Instead, as the rocks, the air, and the sea warm, they emit thermal radiation, or infrared heat. Much of this thermal radiation travels directly out to space, allowing Earth to cool.

Some of this outgoing radiation, however, is reabsorbed by water vapor, carbon dioxide, and the gases in the atmosphere (called greenhouse gases because of their heat-trapping capacity) and is then re-radiated back toward The Earth's surface. On the whole, this re-absorption process is good. If there were no greenhouse gases or clouds in the atmosphere, the Earth's average surface temperature would be a very chilly -18°C (0°F) instead of the comfortable 15°C (59°F) that it is today. What make scientists concerned now is that over the past 250 years, humans have been artificially raising the concentration of greenhouse gases in the atmosphere at an ever-increasing rate. By 2004, humans were pumping out over 8 billion tons of carbon dioxide per year. Some of it was absorbed by "sinks" like forests or the ocean, and the rest accumulated in the atmosphere. We produce millions of pounds of methane by allowing our trash to decompose in landfills and by breeding large herds of methane-belching cattle. Nitrogen-based fertilizers and other soil management practices lead to the release of nitrous oxide into the atmosphere.

[5] explained Policymakers need to know whether prediction is possible and, if so, whether any proposed forecasting method will provide forecasts that are substantially more accurate than those from the relevant benchmark method. An inspection of global temperature data suggests that temperature is subject to irregular variations on all relevant time scales, and that variations during the late 1900s were not unusual. In such a situation, a "no change" extrapolation is an appropriate benchmark forecasting method. We used the UK Met Office Hadley Centre's annual average thermometer data from 1850 through 2007 to examine the performance of the benchmark method. The accuracy of forecasts from the benchmark is such that even perfect forecasts would be unlikely to help policymakers. For example, mean absolute errors for the 20- and 50-year horizons were 0.18C and 0.24C respectively. We nevertheless demonstrate the use of benchmarking with the example of the Intergovernmental Panel on Climate Change's 1992 linear projection of long-term warming at a rate of 0.03 C per year. The small sample of errors from ex ante projections at 0.03 C per year for 1992 through 2008 was practically indistinguishable from the benchmark errors. Validation for long-term forecasting, however, requires a much longer horizon. Again using the IPCC warming rate for our demonstration, we projected the rate successively over a period analogous to that envisaged in their scenario of exponential CO growth—the years 1851 to 1975. The errors from the projections were more than seven times greater than the errors from the benchmark method. Relative errors.

The most obvious impact of global warming will be changes in both average and extreme temperature and precipitation, but warming will also enhance coastal erosion, lengthen the growing season, melt ice caps and glaciers, and alter the range of some infectious diseases, among other things. For most places, global warming will result in more hot days and fewer cool days, with the greatest warming happening over land. Longer, more intense heat waves will become more frequent. High latitudes and generally wet places will tend to receive more rainfall, while tropical regions and generally dry places will probably receive less rain. Increases in rainfall will come in the form of bigger, wetter storms, rather than in the form of more rainy days. In between those larger storms will be longer periods of light or no rain, so the frequency of drought will increase. Hurricanes will likely increase in intensity due to warmer ocean surface temperatures.

[7] discussed about Stream temperature forecasting by means of ensemble of neural networks: Importance of input variables and ensemble size in The Valley of Biala Tarnowska River. Located in the central part of the Polish

Carpathians. The source is located in the Low Beskid at altitudes of 730 meters (Carpathian belt, southern Poland). The total length of the river is 101.8 km and the catchment area to the Koszyce Wielkie gauging station (10 km to the south-west from the city of Tarnow) equals 956.9 km. Biala rnowska catchment is very narrow and extends from the border with Slovakia. The majority of the river has unregulated banks and is in a natural state. Fields, pastures, meadows, and natural vegetation predominate in the catchment of the upper and middle portion of the river. Dominant geology can be defined as sandstone and shale flysch. Biala Tanowska catchment is divided into two different parts. The south section, representing about 25% of the catchment, is a wooded mountain part with the average slope of 10‰. The north part representing almost 75% of the basin, characterized by deep river valleys (mostly agricultural hills and foothills), is generally deforested. The river slope in the northern part is in the range of 0.9–5 ‰. The highest precipitation (up to 100 mm/month), and hence frequent spates are observed during summer months. Average (high/low) temperature is (1ºC/5ºC) in January and (25ºC/13ºC) in July. According to the Köppen Climate Classification Biala Tarnowska river is located within the Humid Continental Zone. It may freeze during winter months and river ice may occur between November and April. If the Biala Tarnowska River is frozen and the results forecasting performance of the neural networks is satisfactory and generally more elements in aggregated ensemble results in better forecast. Prediction errors observed in winter are related to the river freezing and melting processes, which are very difficult to be predicted accurately.

RESEARCH METHODOLOGY

Data and Variable

In this paper, we use daily average temperature data from BMKG, start from January 1st, 2014 to February 7th, 2016.

Analysis Method

There are several steps to analyze. First, because the daily average temperature data { Z_{i} , $i = 1,2,\ldots,600$} ia a time series data, then to model the data using use non-parametric regression B-spline, the data is modified to $\{(x_i , y_i), i = 2,3,...n\} = \{(Z_{i-1} , Z_i), i = 2,3,...n\}$ with n = 600. Second, we use non-parametric regression B-spline by finding the number of knot from orde 2, orde 3, and orde 4. Furthermore, we have to calculate the optimum knot based on GCV minimum in each orde and find prediction based on best model.

RESEARCH RESULT

Using software R, we get knot and GCV values as follows:

Table1. Simulation Using Basis Spline

Orde	Number of Knot	Knot	GCV
Linier*	1	25.21	3,1130
	2*	20.21 , 22.21	3,1082
	3	21.21,22.21 , 23.21	3,1093
Quadratic	1	19.21	3,1295
	2	22.21 , 23.21	3,1222
	3	20.21, 21.21, 22.21	3,1282
Cubic	1	25.21	3,1253
	2	14.21 , 21.21	3,1322
	3	21.21, 22.21, 23.21	3,1394

*Best model

The determination of the best model is based on minimum GCV value. From table 1 above, we get the best model for each orde, which are B-Spline Linier (orde 2) at knot 20,21 and 22,21, B-Spline Quadratic (orde 3) at knot 22,21 and 23,21, and B-Spline Cubic (orde 4) at knot 25,21.

After getting the value of the parameter estimation using linear, quadratic, and cubic models of B-Spline with minimum GCV for each orde, comparison of actual data predictions may be described in figure.1

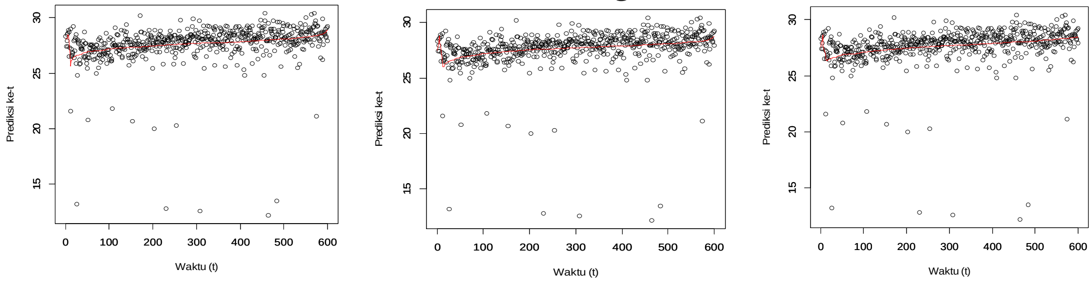

Fig. 1. Results of data actual and prediction based time(t)

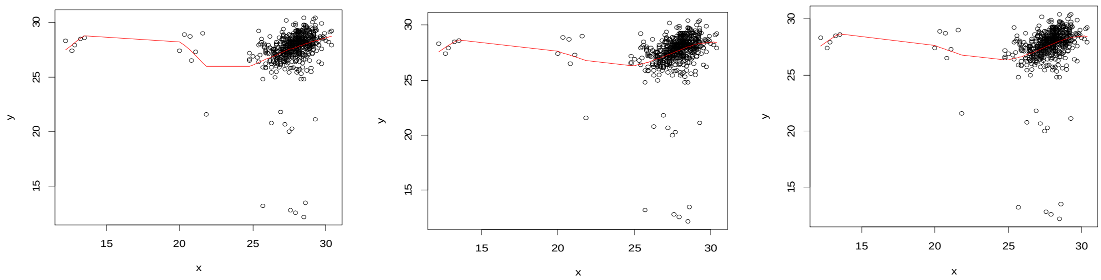

Fig. 2. Results of selection optimal smoothing parameter B-spline

o : actualdata
— : prediction

Curves of B - Spline regression estimation can be seen in Figure 2. The point - black dots in Figure 2 is the actual of temperature data. While the red line represents the value of temperature predictiondatasets $\{(x_i,y_i),\ i=2,3,\ldots n\} = \{(Z_{i-1},Z_i),\ i=2,3,\ldots n\}$ with n = 600.Based on Table 1, we get the best model estimation using B-Spline linier (orde 2) at knots 20,21 and 22,21 with GCV value 3,108242. Therefore, the parameter estimation of B-Spline model is as follows:

$$\hat{y} = 28{,}08631 N_{-1,2}(x) + 28{,}50874 N_{0,2}(x) + 24{,}88\ N_{1,2}(x) + 28{,}89694 N_{2,2}(x)$$

with:

$$N_{-1,2}(x) = \begin{cases} \dfrac{20{,}21 - x}{8{,}02} & ,\quad 12{,}19 \le x \le 20{,}21 \\ 0 & ,\quad others \end{cases}$$

$$N_{0,2}(x) = \begin{cases} \dfrac{x - 12{,}19}{8{,}02} & ,\quad 12{,}19 \le x \le 20{,}21 \\ \dfrac{22{,}21 - x}{2} & ,\quad 20{,}21 \le x \le 22{,}21 \\ 0 & ,\quad others \end{cases}$$

$$N_{1,2}(x) = \begin{cases} \dfrac{x - 20{,}21}{2} & ,\quad 20{,}21 \le x \le 22{,}21 \\ \dfrac{30{,}41 - x}{8{,}2} & ,\quad 22{,}21 \le x \le 30{,}41 \\ 0 & ,\quad others \end{cases}$$

$$N_{2,2}(x)=\begin{cases}\dfrac{x-22{,}21}{8{,}2} & , \quad 22{,}21\le x\le 30{,}41\\ 0 & , \quad \textit{others}\end{cases}$$

We can see more clearly the comparison between forecasting data and actual data on graph as follows **Fig. 3.**

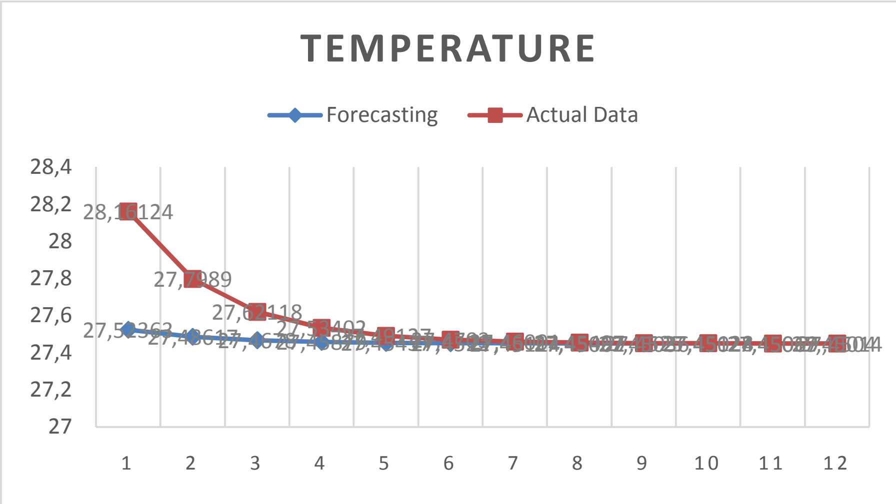

Fig. 3. Comparison Forecasting and Actual Data

The movement of forecasting curve has tendency to follow actual data, so the B-Spline model above can be used to view the indication of the changing in temperature.By using the best model of B-Spline regression, we get forecast value of daily average temperature for the next 12 days, that may be described in Table.2

Table 2. Comparison between Forecasting Data and Actual Data

Date	Forecasting	Actual (Outsample Data)
08-Feb-16	27,52363	27,70
09-Feb-16	27,48617	27,60
10-Feb-16	27,46780	27,70
11-Feb-16	27,45879	27,90
12-Feb-16	27,45437	28,60
13-Feb-16	27,45220	29,00
14-Feb-16	27,45114	28,50
15-Feb-16	27,45062	29,10
16-Feb-16	27,45036	28,70
17-Feb-16	27,45024	28,80
18-Feb-16	27,45017	28,10
19-Feb-16	27,45014	25,60

Based on Table 2, the difference between forecasting data and actual data on February 8th, 2016 is quite small, that is 0,176%. The small differencebetween forecasting data and actual data also can be obtained on February 9th, 2016 - February 12th, 2016, that is under 0,5%. The remaining data have difference between 0,65% - 1,85%

CONCLUSION

Based on the optimal GCV value, the best model of the temperature data obtained using a linear model of B - spline (order2) at 20.21 and 22.21 knots point with GCV value of 3.108242. Based on the graph comparison of actual and predicted based on time (t) using a model B - spline linear (order2), the graph shows that the estimated value equal to the actual value of inflation, the movement of the curve estimation follows the movement of actual inflation, so the regression model B - spline used can used to find indications of rising - falling value of temperature will occur

REFERENCES

[1] Eberly, D., 1999. B-Spline Interpolation on Lattices. *Jurnal* .http://www.e-bookspdf.org, 20 Februari 2014

[2] Filho,W.L. 2009. Communicating climate change:challenges ahead andaction needed. International *Journal of ClimateChange Strategies and Management*, Vol. 1 No. 1, pp. 6-18

[3] H. L. Wei and S. A. Billings. 2006. *An efficient nonlinear cardinal B-spline model for high tide forecastsat the Venice Lagoon*. Nonlinear Processesin Geophysics

[4] Lyche,T., and Mørken,K. 2008, *Spline Methods*, Draft. University of Oslo Ralph J. Cicerone, Climate Change Evidence & Causes. The National Academy of Sciences (NAS)

[5] Hole, S.M, and Teigen, K.H. 2015. Forecasting forecasts: The trend effect. *Judgment and Decision Making*, Vol. 10, No. 5, pp. 416–428

[6] Morisson,J., Quick,M.C., and Foreman,M.G.G. 2002 . Climate Change in The Fraser River watershed: flow and temperature projections. *Journal of Hydrology* 263 pp.230-244

[7] Napiorkowski,M.K. 2014. *Stream temperature forecasting by means of ensemble of neural networks: Importance of input variables and ensemble size.* Taylor & Francis Group, London, ISBN 978-1-138-02674-2.

[8] Nizamitdinov,A., Memmedli,M., and Ozdemir.O. 2010. Comparison Study of P-Spline and Univariate Additive Model (Cubic Smoothing Spline) in Time-Series Prediction. *International conference 24th Mini Euro Conference*

[9] Taylor,J.W., Buizza.R. 2004. A Comparison of Temperature Density Forecasts from GARCH and Atmospheric Models *Journal of Forecasting*, 2004, Vol. 23, pp. 337-355.