The step construction of penalized spline in electrical power load data

Rezzy Eko Caraka*[1], Sakhinah Abu Bakar[2], Gangga Anuraga[3], M A Mauludin[4], Anwardi[5], Suwito Pomalingo[6], Vidila Rosalina[7]
[1,2]School of Mathematical Sciences, Faculty of Science and Technology, University of Malaysia
43600 UKM, Bangi Selangor, +603 8921 4097, Malaysia
[3]Department of Statistics, PGRI Adi Buana Surabaya University, Jl. Dukuh Menanggal XII, Surabaya
60234 Jawa Timur, (031) 8281183, Indonesia
[4]Laboratory of Sociology and Extension, Padjadjaran University
Jl. Raya Bandung-Sumedang Km. 2 Jatinangor, Kab. Sumedang 45363,
Jawa Barat, (022) 842 88898, Indonesia
[5]Industrial Engineering UIN Sultan Syarif Kasim Riau, Jl. Subrantas Km. 15, Pekanbaru, 28293, Indonesia
[6]Informatics Engineering, Universitas Muslim Indonesia, Jl. Urip Sumoharjo KM. 5, Makassar,
Sulawesi Selatan 90231 Indonesia
[7]Faculty of Information Technology, Serang University, Jl. Raya Serang, Cilegon Km 5,
Serang Banten, Indonesia
*Corresponding author, e-mail: Rezzyekocaraka@gmail.com[1], sakhinah@ukm.edu.my[2]

Abstract

Electricity is one of the most pressing needs for human life. Electricity is required not only for lighting but also to carry out activities of daily life related to activities Social and economic community. The problems is currently a limited supply of electricity resulting in an energy crisis. Electrical power is not storable therefore it is a vital need to make a good electricity demand forecast. According to this, we conducted an analysis based on power load. Given a baseline to this research, we applied penalized splines (P-splines) which led to a powerful and applicable smoothing technique. In this paper, we revealed penalized spline degree 1 (linear) with 8 knots is the best model since it has the lowest GCV (Generelized Cross Validation). This model have become a compelling model to predict electric power load evidenced by of Mean Absolute Percentage Error (MAPE=0.013) less than 10%.

Keywords: *forecasting, knot, non-parametric, penalized spline*

1. Introduction

Electricity is a basic need that cannot be separated from humans. Electricity is not only required for lighting, but also as a tool in conducting daily life activities related to socioeconomic society. With the electricity, teaching and learning activities, communication, transportation and health services and development process can run smoothly. So the Government is obliged to provide electricity with sufficient quantity and quality for its people. The availability of power grids in a region can determine progress and developments in the area. However, in reality, the provision of electrical energy by PT PLN (Persero) [1] as the official institution appointed by the government to manage the electricity problem in Indonesia, has not been able to meet the people's need for electrical energy as a whole. The geographical condition of the Indonesian state, which consists of thousands of islands dispersed and unevenly buried power centers. Weak electricity demand in some areas, the high cost of marginal development of electricity supply system and limited financial ability are inhibiting factors of energy supply Electricity on a national scale [2].

Given the importance of the availability of electric energy in an area, a precise method is needed to assess the total use of it in the future [3]. Daily total electricity usage is time series data, but the data does not meet stationary and white noise assumption so that the parametric regression is less appropriate to use [4]. To overcome this, we use nonparametric regression that does not require any assumptions [5]. One of the nonparametric regressions that can be used to model data is Spline Regression [6]. Spline regression is one approach to the commonly used nonparametric method because it gives excellent flexibility [7] to the

Received January 18, 2018; Revised October 23, 2018; Accepted November 10, 2018

characteristics of a function or data, and is capable of handling data character or function that is smooth. In spline regression, the best model determination by choosing the exact number and location of the knots [8]. This process usually takes a long time and if done using the software also requires a large memory [9]. To overcome this, we can use Penalized spline regression (P-Spline) [10] where this method utilizes the quantitative points of the unique (single) value of the predictor variable as the place of the knot [11] so as to produce better flexibility [12]. By using penalized spline regression, we will predict the total daily usage data in Sumatra. Predictive data is expected to be used as a reference for the parties concerned in determining the direction of the policy of the fulfillment of electrical energy is appropriate and following the request.

2. Research Method

2.1 Non Parametric Regression

Smoothing is one of the methods used in nonparametric data analysis [13]. The purpose of smoothing is to minimize the diversity and estimate the behavior of data that tends to be different and has no effect so that the characteristics of the data will appear more clearly [14]. Spline popular in Ecology [4] and biodiversity [15] One of the regression models with a nonparametric approach that can be used to estimate the regression curve is spline regression[16]. Spline regression is an approach to matching data while still taking into account the smoothing curve. The spline approach has its virtue because spline is a piecewise polynomial of order m that has a continuous segmented property that adequately describes the local characteristics of data functions [6] Spline regression with the orde of m and the knot point $\tau_{1,}\ \tau_2$, ..., τ_K can be explained in (1):

$$y_i = \beta_0 + \beta_1 x + \cdots + \beta_{m-1}x^{m-1} + \sum_{k=1}^{K}\beta_{m-1+k}(x-\tau_k)_+^{m-1} + \varepsilon_i \qquad (1)$$

with truncated function

$$(x-\tau_k)_+^{m-1} = \begin{cases}(x-\tau_k)^{m-1}, & x \geq \tau_k \\ 0 & ,\ x < \tau_k\end{cases} \qquad (2)$$

Moreover, m as polynomial orde, τ_k as knot point to-k with k=1, 2,.., K and ε_i as error independent random assumed to be normal with mean zero and variance σ^2 [17]. Spline has an advantage in overcoming the pattern of data showing sharp ups and downs with the help of knots, and the resulting curve is relatively smooth. Knots are a common fusion that shows the behavioral changes of the spline function at different intervals [18]. One form of spline regression is the penalized spline obtained by minimizing Penalized Least Square (PLS), which is an estimation function that combines the least square function and smoothness of the curve. Penalized spline regression [19] is a popular smoothing approach [12] the penalized spline consists of the polynomial piecewise [20] that have a continuous segmental property [21]. This property provides better flexibility than ordinary polynomials making it possible to adopt functional or data characteristics efficiently.

The polynomial basis function in the penalized spline estimator is less capable of handling an arbitrary data and numerical instability when the number of large knot points and the smoothing parameter value (λ) is small or 0. To treat an arbitrary data and uncertainty [22], numerical then change the function of a base polynomial on penalized spline estimator with radial base function [23]. The radial base function is a function that depends on the distance between the data with a data center. In regression, the penalized spline is used radial basis function with K knot, for example, $\tau_1, \tau_2,..., \tau_K$ can be explain $1, x,..., x^{m-1}, |x-\tau_1|^{2m-1}, |x-\tau_2|^{2m-1}..., , |x-\tau_K|^{2m-1}$. *W*here m is a polynomial order and τ_k is the kth knot with k = 1, 2, .., 3 which is the data center. Given n paired observations $(x_1,y_1),\ (x_2,y_2),...,(x_n,y_n)$ following the non-parametric regression model in (3):

$$y_i = f(x_i) + \varepsilon_i \qquad (3)$$

with ε_i is a random error assumed to be independent with mean 0 and variance σ^2 and $f(x_i)$ is an unknown form of the regression function. The function $f(x_i)$ using radial-based penalized splines in (4):

$$f(x_i)=\sum_{r=0}^{m-1}\beta_r\ x_i^r+\sum_{k=1}^{K}\beta_{mk}\ |x-\tau_k|^{2m-1}+\varepsilon_i \text{ for m=2,3...} \tag{4}$$

if the equation is elaborated then obtained the following translation system:

$$\begin{aligned}
f(x_1)&=\beta_0+\beta_1 x_1+\ldots+\beta_{m-1}x_1^{m-1}+\beta_{m1}\,|x_1-\tau_1|^{2m-1}+\ldots+\beta_{mK}\,|x_1-\tau_K|^{2m-1}+\varepsilon_1\\
f(x_2)&=\beta_0+\beta_1 x_2+\ldots+\beta_{m-1}x_2^{m-1}+\beta_{m2}\,|x_2-\tau_2|^{2m-1}+\ldots+\beta_{mK}\,|x_2-\tau_K|^{2m-1}+\varepsilon_2\\
&\vdots\\
f(x_n)&=\beta_0+\beta_1\,x_n+\ldots+\beta_{m-1}\,x_n^{m-1}+\beta_{mn}\,|x_n-\tau_n|^{2m-1}+\ldots+\beta_{mK}\,|x_n-\tau_K|^{2m-1}+\varepsilon_n
\end{aligned} \tag{5}$$

shown in (5) can be written in matrix form:

$$\begin{bmatrix} f(x_1)\\ f(x_2)\\ \vdots\\ f(x_n)\end{bmatrix}=\begin{bmatrix}1 & x_1 & x_1^2 & \cdots & x_1^{m-1} & |x_1-\tau_1|^{2m-1} & \cdots & |x_1-\tau_K|^{2m-1}\\ 1 & x_2 & x_2^2 & \cdots & x_2^{m-1} & |x_1-\tau_1|^{2m-1} & \cdots & |x_2-\tau_K|^{2m-1}\\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots\\ 1 & x_n & x_n^2 & \cdots & x_n^{m-1} & |x_1-\tau_1|^{2m-1} & \cdots & |x_n-\tau_K|^{2m-1}\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_{m-1}\\ \beta_{m1}\\ \beta_{m2}\\ \vdots\\ \beta_{mK}\end{bmatrix}+\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix}$$

or

$$f(x)=\mathbf{X\beta}+\boldsymbol{\varepsilon} \tag{6}$$

with

$$f(x)=(f(x1), f(x\ 2),\ldots, f(xn))T, \beta=(\beta0, \beta1,\ldots, \beta m\text{-}1, \beta m1,\ldots, \beta mK)T$$

and

$$\boldsymbol{X}=\begin{bmatrix}1 & x_1 & x_1^2 & \cdots & x_1^{m-1} & |x_1-\tau_1|^{2m-1} & \cdots & |x_1-\tau_K|^{2m-1}\\ 1 & x_2 & x_2^2 & \cdots & x_2^{m-1} & |x_1-\tau_1|^{2m-1} & \cdots & |x_2-\tau_K|^{2m-1}\\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots\\ 1 & x_n & x_n^2 & \cdots & x_n^{m-1} & |x_1-\tau_1|^{2m-1} & \cdots & |x_n-\tau_K|^{2m-1}\end{bmatrix}$$

The estimation result of regression function *f*(*x*) in (6) is:

$$\hat{f}\,(x)=\boldsymbol{X}\,\hat{\boldsymbol{\beta}} \tag{7}$$

The radial-based penalized spline regression estimation is obtained by minimizing Penalized Least Square (PLS). Penalized Least Square (PLS) is a function of estimation criteria that combines least square functions with smoothness sizes. PLS functions can be describe in (8):

$$Q=\ \|\boldsymbol{y}-\boldsymbol{X\beta}\|^2+\lambda^{2m-1}\boldsymbol{\beta}^T\boldsymbol{R\beta},\ \ \lambda>0 \tag{8}$$

where λ as smoothing parameter , R= $\begin{bmatrix}0_{mxm} & 0_{mxK}\\ 0_{Kxm} & P_{KxK}\end{bmatrix}$,

$\mathbf{P}=[P_{ab}]$ with $P_{ab}=\ |\tau_a-\tau_b|^{2m-1}\,for\ 1\le a\le K\ and\ 1\le b\le K$

$\boldsymbol{\beta}^{\mathbf{T}}=(\beta_0, \beta_1,.., \beta_{m\text{-}1}, \beta_{m1},..., \beta_{mK})$

shown in (8) can be describe:

$$Q = \|y - X\beta\|^2 + \lambda^{2m-1}\beta^T R\beta$$

$$Q = (Y - X\beta)^T (Y - X\beta) + \lambda^{2m-1}\beta^T R\beta$$

$$= (Y^T - \beta^T X^T)(Y - X\beta) + \lambda^{2m-1}\beta^T R\beta$$
$$= (Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta) + \lambda^{2m-1}\beta^T R\beta$$

$$= (Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta) + \lambda^{2m-1}\beta^T R\beta$$

Parameter of $\boldsymbol{\hat{\beta}}$ from (7) obtained by minimizing PLS function. Sufficient condition for PLS function to reach minimum value is $\frac{\partial Q}{\partial \boldsymbol{\beta}} = 0$ So obtained

$$\frac{\partial Q}{\partial \beta} = (0 - 2X^T Y + 2X^T X\beta) + 2\lambda^{2m-1}R\beta = 0$$

$$X^T X\beta + \lambda^{2m-1}R\beta = X^T Y$$

$$\beta\,(X^T X + \lambda^{2m-1}R) = X^T Y$$

$$\hat{\beta} = (X^T X + \lambda^{2m-1}R)^{-1}X^T Y \qquad (9)$$

substitute (9) to (7) to obtain an estimate of *f*(*x*):

$$\hat{f}\,(x) = X(X^T X + \lambda^{2m-1}R)^{-1}X^T Y \qquad (10)$$

shown on (10) can be expressed in form:

$$\hat{f}\,(x) = S(\lambda)\,Y \qquad (11)$$

with $\boldsymbol{S(\lambda) = X(X^T X + \lambda^{2m-1}R)^{-1}X^T}$ therefore the estimation of nonparametric regression function $\boldsymbol{\hat{f}\,(x)}$ on (11) depends on the parameters of the λ.

2.2 Smoothing Parameter and Optimum Knots

The smoothing parameter as well as the balance controller between the curve conformity to actual data and training data. Pairing very small or large smoothing parameters will provide a very coarse or smooth form of completion function [6]. On the other hand, it is desirable to have an estimator form in addition to having a degree of the knot, also in accordance with the data [17]. Therefore, it is essential to choose an optimal finishing parameter. Selecting a smoothing parameter in principle is equivalent to selecting the optimal number of knots that produce the optimal knot value [24] which results in the minimum GCV value [25]. The GCV (Generalized Cross Validation) function can be expressed in equation (12):

$$GCV = n^{-1}\frac{RSS(\lambda)}{(1-n^{-1}df)^2} \qquad (12)$$

with RSS = *Residual Sum Square*, $RSS(\lambda) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$. According to [7], the degrees of freedom are equivalent to the trace values of the hat matrix

$$S(\lambda) = (X(X^T X + \lambda^{2m-1}R)^{-1}X^T)$$

$$df = tr(S(\lambda)) = tr(X(X^T X + \lambda^{2m-1}R)^{-1}X^T)$$

so, the GCV function can be explained in (13)

$$GCV = n^{-1}\frac{RSS(\lambda)}{(1-n^{-1}tr(S(\lambda)))^2} \qquad (13)$$

3. Results and Analysis

First, we must modify the time series data into two variables so that it can be used to do data estimation by regression approach through two variables. The two variables are independent variables (last day's data) and the dependent variable (the value to be predicted on this day). The characteristics of the smoothing regression are analyzed within the framework of identical and independent observation structures. Assume a pair of $\{X_i, Y_i\}_{i=1}^{n}$ is independent. In reality the situation is inconsistent with the assumption that the observations (X_1,Y_1), (X_2,Y_2),...., (X_n,Y_n) are independent. Thus, if an object is observed from time to time, it is very likely that the dependent object will be affected by the dependent of the previous object. These effects can be modeled in three forms:

a. Model (S): a stationary sequence {(X_1, Y_1), i = 1,2,...,n} is the result of observation and will be estimated *f*(*x*) = E(Y|X=x)
b. Model (T): a time series {Z_i, i ≥ 1} is the result of observation and in used to predict Z_{n+1} with *f*(*x*) = E(Y|X=x)
c. Model (C): error observation {ε_{in}} in the regression model forms the row correlated random variables

As we know characteristics of electrical data is complicated. Therefore, to predict the time series problem (T) one dimension can be drawn to the first model. By setting the stationary time series {Z_i, i ≥1}. The lag value of Z_{i-1} as X_i and Z_i as Y_i. Then for the problem of estimating Z_{n+1} from $\{Z_i\}_{i=2}^{n}$ can be considered as smoothing regression problem of $\{X_i, Y_i\}_{i=2}^{n} = \{Z_{i-1}, Z_i\}_{i=2}^{n}$. The prediction problem for time series {*Zi*} is the same as the estimation of *f*(*x*) = E(Y|X=x) for two time series dimensions $\{X_i, Y_i\}_{i=1}^{n}$. In this paper, we provide the step construction of electrical load data by using p-spline radial basis as follows:

1. Modify the form of time series data into two variables dependent variable and independent variable. Z_t or current data is the dependent variable (Y_i) whereas Z_{t-1} or the previous day's data is an independent variable (X_i).
2. Define a new independent variable X containing the unique value of the independent variable X that has been sorted from the smallest to the largest value.
3. Determining the order, many knots and the value of the fining parameters.
4. Determine the knot points using a quantitative sample of the new independent variable X. Where the quantitative sample is based on the many knots (K) used.
5. Calculate parameter of $\hat{\boldsymbol{\beta}}$ according to the equation:

$$\hat{\beta} = (X^T X + \lambda^{2m-1} R)^{-1} X^T Y \tag{14}$$

where

$$R = \begin{bmatrix} 0_{mxm} & 0_{mxK} \\ 0_{Kxm} & P_{KxK} \end{bmatrix},\ P = [P_{ab}] \tag{15}$$

when

$$P_{ab} = |\tau_a - \tau_b|^{2m-1} \text{ for } 1 \le a \le K \text{ } and \text{ } 1 \le b \le K \tag{16}$$

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{m-1} & |x_1 - \tau_1|^{2m-1} & \cdots & |x_1 - \tau_K|^{2m-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{m-1} & |x_1 - \tau_1|^{2m-1} & \cdots & |x_2 - \tau_K|^{2m-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{m-1} & |x_1 - \tau_1|^{2m-1} & \cdots & |x_n - \tau_K|^{2m-1} \end{bmatrix} \tag{17}$$

6. Estimate the model according to the (18):

$$\hat{f}\left(x; \hat{\beta}\right) = X(X^T X + \lambda^{2m-1} R)^{-1} X^T Y = X\,\hat{\beta} \tag{18}$$

7. Calculate GCV as in (12) and (13) and select the optimal smoothing parameters from each of the many knot points based on minimum GCV values.

8. Selecting many optimal knots from each order using full-search algorithm method.
9. Estimate the model of each order based on many optimal knot points and optimal finishing parameters.
10. Choosing an optimal model based on minimum GCV values.
11. Calculate the value of R^2 to find out how far the match of a model and calculate the MAPE value see [26] more details, to determine the percentage of error between the actual data and predicted data.
12. Compare the predicted results with the actual data available

This study uses secondary data obtained from BP3Sumatera. The data used is the total daily electricity usage data in Sumatra starting from the period of 1 August 2015 to 31 December 2015. The data is divided into two that are *in sample* data which is used to construct the model (starting from August 1st, 2015 to November 30th, 2015) and *out sample* data which are used to determine the accuracy of the model (starting from December 1st, 2015 to December 31st, 2015). The variables used in this study were the total daily usage of electricity in Sumatra where the predictor variable (Xi) is the total data of electricity usage at time 1.2, ..., i-1 k while the response variable (Yi) is the lag 1 of the total data Use of electricity i.e. data to 2,3,…,i. The average daily usage of electricity in Sumatra is 109162W with standard deviation of 2758W. The lowest utilization of electricity in Sumatra in August to November 2015 occurs on Sunday, August 2nd, 2015 which is 101313W. This is thought to be due to an earthquake of 4.7 SR in the Mentawai, Padang to Painan, which caused the breaking of electricity for some time in the Mentawai and surrounding areas. While the largest use of electricity occurred on Tuesday, 27 October 2015.

3.1. Modelling of Penalized Spline Regression with Degree 1

By using knots as much as 20, it is obtained that the optimum parameter of λ is 681 with minimum GCV of 4635405. Meanwhile, the optimal knots are 18 knots, so the formed model is:

$$\begin{aligned} f(x) = {} & -3189.044 + 1.04162x + 2.420953\,(x - 104036.3) - 2.975856(x + 104934.5) \\ & - 0.6093703(x + 105764.9) - 8.489145(x + 106981.9) + 10.16902(x \\ & + 107352.1)\ 11.49861(x + 108273.7) - 12.89382(x + 108485.2) \\ & - 1.882402(x + 109125.5)\ 5.305015(x + 109290.7) - 3.197387(x109783.9) \\ & - 17.73622(x + 110371.1)\ 23.47804(x + 110558.3) - 3.678277(x \\ & + 110898.8) - 2.278078(x + 111253.5)\ 0.8062679(x + 111620.3) \\ & - 3.569115(x + 111779.7) + 2.80086(x + 112256.9)\ 3.949552(x + 112714.2) \end{aligned}$$

3.2. Modelling of Penalized Spline Regression with Degree 2

By using knots as much as 20 it is obtained that the parameter of λ optimizer is 661 with minimum GCV of 4474538. Meanwhile, the optimal knots are 8 knots, so the formed model is:

$$\begin{aligned} f(x) = {} & 187155.4 - 3.307654\mathrm{x} + 0.0002.430867\,\mathrm{x2} - 0.0009850908(\mathrm{x} \\ & + 104942.7)\ 0.002900279(\mathrm{x} + 107018.7) - 0.002841148(\mathrm{x} + 108361) \\ & - 0.0003218168(\mathrm{x} + 109209) + 0.00257259(\mathrm{x} + 110114.3) - 0.001327075(\mathrm{x} \\ & + 110602) - 0.001098519(\mathrm{x} + 111569.3) + 0.003364731(\mathrm{x} + 112240.7) \end{aligned}$$

3.3. Modeling of Penalized Spline Regression with Degree 3

By using knots as much as 20 it is obtained that the parameters of the optimized λ membalus of 7290 with a minimum GCV of 13150140. Meanwhile, the optimal knots amounted to 3 knots. So that the model is formed:

$$\begin{aligned} f(x) = {} & 84690.47 + 1.019746\mathrm{e} - 14\mathrm{x} - 4.067427\mathrm{e} - 14\mathrm{x2} + 1.879564\mathrm{e} - 11\,\mathrm{x3} - 1.822854\mathrm{e} \\ & - 28\,(\mathrm{x} + 107251.2) - 1.925563\mathrm{e} - 28(\mathrm{x} + 109498) - 1.32593\mathrm{e}28(\mathrm{x} \\ & + 111308.8) \end{aligned}$$

3.4. Selection of the Best Penalized Spline Regression Model

Selection of the best-penalized spline regression model is done by selecting the smallest Optimal GCV value. Based on Table 1, the optimum minimum GCV value is at 1st degree of 103300 with 18-point knots and 681 optimal λ. The best regal penalized spline model is:

$$\begin{aligned}f(x) = -3189.044 &+ 1.04162\mathrm{x} + 2.420953\,(\mathrm{x} - 104036.3) - 2.975856(\mathrm{x} + 104934.5)\\ &- 0.6093703(\mathrm{x} + 105764.9) - 8.489145(\mathrm{x} + 106981.9) + 10.16902(\mathrm{x}\\ &+ 107352.1)\ 11.49861(\mathrm{x} + 108273.7) - 12.89382(\mathrm{x} + 108485.2) - 1.882402(\mathrm{x}\\ &+ 109125.5)\ 5.305015(\mathrm{x} + 109290.7) - 3.197387(\mathrm{x}109783.9) - 17.73622(\mathrm{x}\\ &+ 110371.1)\ 23.47804(\mathrm{x} + 110558.3) - 3.678277(\mathrm{x} + 110898.8) - 2.278078(\mathrm{x}\\ &+ 111253.5)\ 0.8062679(\mathrm{x} + 111620.3) - 3.569115(\mathrm{x} + 111779.7) + 2.80086(\mathrm{x}\\ &+ 112256.9)\ 3.949552(\mathrm{x} + 112714.2)\end{aligned}$$

Table 1. The Comparison of GCV Value

Degree	λ Optimal	Number of knot	Knot Point	Optimal GCV
1	681	18	104036.3 104934.5 105764.9 106981.9 107352.1 108273.7 108485.2 109125.5 109290.7 109783.9 110371.1 110558.3 110898.8 111253.5 111620.3 111779.7 112256.9 112714.2	103300
2	661	8	104942.7 107018.7 108361 109209 110114.3 110602 111569.3 112240.7	4474538
3	7290	3	107251.2 109498 111308.8	13150140

Based on the data out sample obtained MAPE value of 0.0131523. The value <10% means that the model's performance is very accurate [27]. Daily total daily electricity usage prediction for the period of 10-31 December 2015 is made by using penalized spline regression model with degree 1. Since the total daily usage value in Sumatera is already known, it can be done comparison between the actual value and the prediction value according to Table 1 and Figure 1

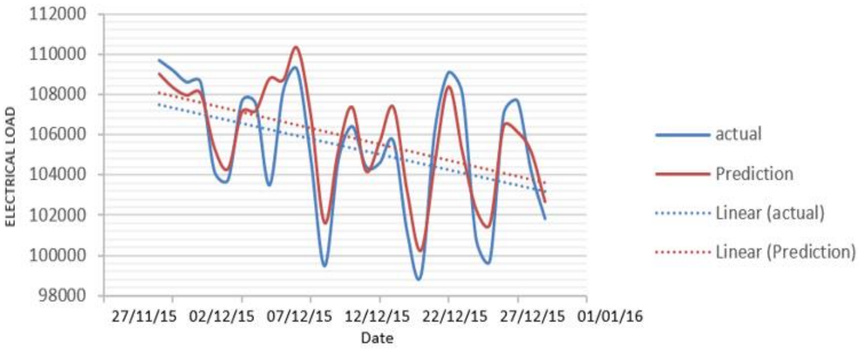

Figure 1. Actual data and prediction data

4. Conclusion

Nonparametric approaches to estimation of regression curves have flexibility. The flexibility of the method is advantageous in research to find a particular relation of a case that is not or has not been previously available information on the shape of the regression curve. In fact, the regression curves obtained with nonparametric techniques may provide suggestions in constructing appropriate parametric models for future studies The optimum parameter of λ is obtained with minimum Generalized Cross-Validation (GCV) criterion, while optimal knot selection is done by full-search algorithm method. Generally, the knots of a P-spline are at xed quantiles of the independent variable and the only tuning parameter to choose is the number of knots. In this article, the effects of the number of knots on the performance of P-splines are studied. Two algorithms are proposed for the automatic selection of the number of knots. The myoptic algorithm stops when no improvement in the generalized cross validation statistic (GCV) is noticed with the last increase in the number of knots. The full search examines all candidates in a x-ed sequence of possible numbers of knots and chooses the candidate that minimizes GCV. The myoptic algorithm works well in many cases but can stop prematurely. The full search algorithm worked well in all examples examined. Based on this research we have successfully done the simulation of soft computing using penalized spline and got a conclusion that this model can be used for electrical data which is known acutely fluctuate

Acknowledgment

This research fully supported by GUP-2017-118 Universiti Kebangsaan Malaysia. This research also under collaboration Department of Statistics, Universitas PGRI Adi Buana Surabaya, Universitas Padjadjaran, Industrial Engineering UIN Sultan Syarif Kasim Riau, Informatics Engineering, Universitas Muslim Indonesia, and Faculty of Information Technology, Universitas Serang Raya.

References

[1] Caraka RE, Ekacitta PC. Simulation of the New Renewable Energy Calculator to Meet Energy Security in Indonesia (in Indonesia: Simulasi Kalkulator Energi Baru Terbarukan (EBT) Guna Memenuhi Ketahanan Energi di Indonesia). *Statistika*. 2016; 16(2): 77-78.

[2] Ministry of Energy and Mineral Resources Indonesia. Official Electricity Statistics 2017 (in Indonesia: Statistik Ketenagalistrikan 2017). Dir. Gen. Electr. Energy Util. 2017.

[3] Torío HDS, Angelotti A. Exergy analysis of renewable energy-based climatisation systems for buildings. a Critical View. *Energy and Buildings*. 2009; 41(3): 248–271.

[4] Caraka RE, et al. Ecological Show Cave and Wild Cave : Negative Binomial Gllvm's Arthropod Community Modelling. *Procedia Computer Science*. 2018; 135: 377–384.

[5] Hollingsworth B. Non-parametric and parametric applications measuring efficiency in health care. Health *Care Management Science*. 2003; 6(4): 203-218.

[6] Eubank RL. A simple smoothing spline, III. *Computational Statistic*.2004; 19(2): 227–241.

[7] Ruppert D. Wand MP, Carrol RJ. Semiparametric regression during 2003–2007*. Electron. J. Stat*. 2009; 3: 1193–1256.

[8] de Boor C. Spline basics. Handbook of Computer Aided Geometric Design. 2002.

[9] Sutiksno DU. Gio PU, Caraka RE, Ahmar AS. Brief Overview of STATCAL Statistical Application Program. *Journal of Physics: Conference Series*. 2018; 1028(012244): 1-13.

[10] Claeskens G, Krivobokovab T, Opsomer JD. Asymptotic properties of penalized spline estimators. *Biometrika*. 2009; 96(3): 529–544.

[11] Caraka RE. Supatmanto BD, Soebagyo J, Maulidin MA, Iskandar A, Pardamean B. Rainfall Forecasting Using PSPLINE and Rice Production with Ocean-Atmosphere Interaction. *IOP Conference Series Earth and Environmental Science*. 2018.

[12] Wang X, Shen J, Ruppert D. On the asymptotics of penalized spline smoothing. *Electronic Journal of Statistic*. 2011; 5: 1–17.

[13] Caraka RE, Bakar SA. Evaluation Performance of Hybrid Localized Multi Kernel SVR (LMKSVR) In Electrical Load Data Using 4 Different Optimizations. *Journal of Engineering and Applied Science*. 2018; 13(17): 7440-7449.

[14] Prahutama A, Utami TW, Caraka RE, Zumrohtuliyosi D. Pemodelan Inflasi Berdasarkan Harga-Harga Pangan Menggunakan Spline Multivariabel. *Media Statatistika*. 2014; 7(2): 89–94.

[15] Kurniawan ID, Rahmadi C, Caraka RE, Ardi TA. Short Communication: Cave-dwelling Arthropod community of Semedi Show Cave in Gunungsewu Karst Area, Pacitan, East Java, Indonesia. *Biodiversitas*. 2018; 19(3): 857–866.

[16] Crawley MJ. The R Book. John Wiley and Sons. 2012.

[17] Ruppert D. Nonparametric Regression and Spline Smoothing. *Journal of the American Statistical Association*. 2001; 96(456): 1522-1523.

[18] Caraka RE, Sugiyarto W. Inflation Rate Modelling In Indonesia. *Etikonomi*. 2016; 15(2): 111–124.

[19] Yu Y, Wu C, Zhang Y. Penalised spline estimation for generalised partially linear single-index models. *Statistics Computing*. 2017; 27(2): 571–582.

[20] Krivobokova T, Kneib T, Claeskens G. Simultaneous confidence bands for penalized spline estimators. *Journal of the American Statistical Association*. 2010; 105(490): 852–863.

[21] Pütz P, Kneib T. A penalized spline estimator for fixed effects panel data models. *AStA Advances in Statistical Analysis*.2017; 102(4): 1–22.

[22] Suparti, Caraka RE, Warsito B, Yasin H. The Shift Invariant Discrete Wavelet Transform (SIDWT) with Inflation Time Series Application. Journal of Mathematics. Research. 2016; 8(4): 14–20.

[23] Caraka RE, Yasin H, Basyiruddin AW. Forecasting Crude Palm Oil (CPO) Using the Vector Regression Kernel Base Radial Support (in Indonesia: Peramalan Crude Palm Oil (CPO) Menggunakan Support Vector Regression Kernel Radial Basis). *Jurnal Matematika Universitas Udayana*. 2017; 7(1): 43–57.

[24] Caraka RE, Devi AR. Application of Non Parametric Basis Spline (B-Spline) in Temperature. *Statistika*. 2016; 4(2): 68–74.

[25] Lukas MA, de Hoog FR, Anderssen R.S.Practical use of robust GCV and modified GCV for spline smoothing. *Computational Statatics*. 2016; 31(1): 269–289.

[26] GEP, Jenkins GM, Reinsel GC/ Time Series Analysis: Forecasting & Control. 1994.

[27] Subba R T. Time Series Analysis. *J. Time Ser. Anal*. 2010; 31(2): 139.