© 2019 IJSRSET | Volume 6 | Issue 1 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology
DOI : https://doi.org/10.32628/IJSRSET211841125

Truncated Spline Regression to Estimate Curve of Strontium Titanate XRD Data

Bramadita[1], Aji Hamim Wigena[2], Muhammad Nur Aidi[3]
[1,2,3]Department of Statistics, Bogor Agricultural University, Bogor, West Java, Indonesia
bramadita_22@yahoo.com

ABSTRACT

Characterization of ferroelectric material using X-Ray Diffraction tools (XRD) resulted in 2θ degree angle and intensity of diffraction data. These data showed the relationship between 2θ degree angle and diffraction intensity which formed a spectrum pattern with fluctuation of diffraction intensity values as the increasing of 2θ degree angles. The high fluctuation of diffraction intensity value affected the inconsistency of mean and variance values. The estimation of the spectrum pattern could not be analysed by parametric models which needs strict with assumptions. Spline regression is a polynomial model with flexible segmentation to estimate the curve of XRD data. This research used the truncated spline regression model to estimate the curve of Strontium Titanate ($SrTiO_3$) and $SrTiO_3$ doping with 2% of RuO_2 ($SrTiO_3$+RuO_2) XRD data. The best curve of $SrTiO_3$ was estimated by the model with 42 knots, while that of $SrTiO_3$+RuO_2 was estimated by the model with 37 knots.

Keywords : Knots, Strontium Titanate, Truncated Spline Regression, X-ray Diffraction.

I. INTRODUCTION

The thin films of Strontium Titanate ($SrTiO_3$) and $SrTiO_3$ doping with 2% Ruthenium Oxide ($SrTiO_3$+RuO_2) was successfully grown by the Department of Physics, Bogor Agricultural University. These films were then characterized using X-Ray Diffraction (XRD) tools. The Characterization resulted in 2θ degree angle and intensity of diffraction data. Each peak that appears in the XRD pattern represents a crystal field representing a particular orientation. The peaks on the XRD pattern were then compared to the ICDD (International Center for Diffraction Data) to determine the crystal structure of a material (Widyawati 2012).

The use of statistical models for analyzing XRD data is very potential to be used as further material analysis in physics. One of the models is regression to estimate a curve of XRD data in order to find the actual peaks of the material diffraction intensity. This is useful for light, heat and humidty sensors.

The result of XRD characterization showed that the relationship between 2θ degree angle and diffraction intensity formed spectrum pattern with fluctuation of diffraction intensity values as the increasing of 2θ degree angles. The high fluctuation of diffraction intensity value affected the inconsistency of mean and variance values. The estimation of the spectrum pattern could not be analysed by parametric models which needs strict with assumptions. Therefore, it is necessary to form a model to predict the shape of the curve through a nonparametric approach. One of the models with a nonparametric approach that is often used to estimate curve shape is spline regression.

Spline regression is a method used to obtain the estimated regression curve through the data fitting method. This method is a polynomial model with segmented character that has high flexibility. These properties allow the spline regression model to be adapted to the local characteristics of the data. The advantage of the model with the spline approach is that the model formed can consider data patterns that rise or fall sharply with the help of point knots, and the resulting curve is relatively smooth (Hardle 1990). Knot points are the meeting points of spline behavior patterns at different intervals (Tripena 2011). Optimal knots are determined based on the smallest General Cross Validation (GCV) value (Suparti et al. 2013).

In this study the XRD data of $SrTiO_3$ and $SrTiO_3$+RuO_2 materials were modeled using truncated spline regression analysis to estimate the regression curve between *2θ* degree angle and diffraction intensity. The model used were a linear and quadratic model. The criteria used to choose the best spline regression model is a model that has optimal knots with the smallest Akaike's Information Criterion (AIC) value and the largest determination coefficient (R^2) between the two models.

II. METHODS AND MATERIAL

The data in this study is from the Department of Physics, Bogor Agricultural University. These data are XRD data of $SrTiO_3$ and $SrTiO_3$+RuO_2. The response variables are the diffraction intensity of $SrTiO_3$ (Y_1) and $SrTiO_3$+RuO_2 (Y_2) materials. The explanatory variable is the *2θ* degree angle (X).

The steps of the analysis are described as follows:

1. Data exploration to see the shape of the curve or data pattern;
2. Selecting the number and location of knots. The selection of the knot points uses the Full Search Algorithm. Knot points are selected from all data on the explanatory variable as a candidate knots, then computes the GCV value. The knot points that minimize GCV is chosen to be knots used (Ruppert 2000).
3. Selecting the number and location of the optimum knot points based on the smallest GCV value with the following equation (Spriti et al. 2012) :

$$\mathrm{GCV} = \frac{\mathrm{MSE}}{(\mathrm{n}^{-1}\mathrm{tr}(\mathbf{I} - \mathbf{A}(\mathbf{k})))^2} \quad (1)$$

with MSE is the mean squared error, n is the number of observations, I is the identity matrix and $\mathbf{A(k)}$ is $\mathbf{X(X^TX)^{-1}X}$.

4. Estimating the truncated spline regression parameter using the least squares method. The estimator of truncated regression parameters is obtained by the following equation:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X^TVX})^{-1}\mathbf{X^TVY} \quad (2)$$

5. Forming a truncated spline regression model based on the parameters obtained in the third and fourth step. The truncated polynomial of degree *D* with a knot *k* is the function which is equal to 0 to the left of *k* and equal to $(x-k)^D$ to the right of *k*.

$$(\mathrm{x}-\mathrm{k})_+^{\mathrm{D}} = \begin{cases} 0, & \mathrm{x} < \mathrm{k} \\ (\mathrm{x}-\mathrm{k})^{\mathrm{D}}, & \mathrm{x} \geq \mathrm{k} \end{cases} \quad (3)$$

The equation for a spline of degree *D* with *K* knots is (Eubank 1999) :

$$\mathrm{y} = \beta_0 + \sum_{d=1}^{D} \beta_d x^d + \sum_{k=1}^{K} b_k (x-k)_+^D \quad (4)$$

with β_0 is intecept, β_d is regression parameter in the d-th degree and b_k is regression parameter in the k-th knots.

6. Selecting the best spline truncated regression model based on the smallest AIC value and the largest R^2 with the following equation:

$$\mathrm{AIC} = \mathrm{Ln(MSE)} + 2\frac{\mathrm{P}}{\mathrm{n}} \quad (5)$$

with P is the number of parameters estimated and n is the number of observations. In addition, the model suitability measure can be seen from the R^2 by the following equation:

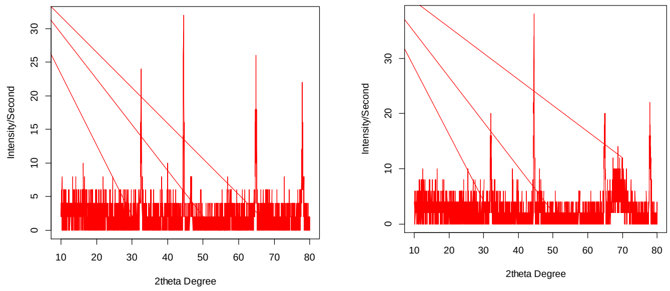
$$R^2 = \frac{\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \qquad (6)$$

with $\widehat{Y}_i$ is the estimated value of the response in the i-th observation, Y_i is the actual value of the i-response and $\overline{Y}$ is the average value of the actual response.

Steps 2 through 5 are carried out for linear and quadratic truncated spline regression models.

III. RESULTS AND DISCUSSION

Characterization using XRD equipment resulted a line diagram between the *2θ* degree angle and diffraction intensity. Figure 1 shows a line diagram between the variable X and the variable Y on each material.



(a) $SrTiO_3$ (b) $SrTiO_3+RuO_2$

Figure 1 : Line Diagram Between *2θ* degree angle and Diffraction Intensity

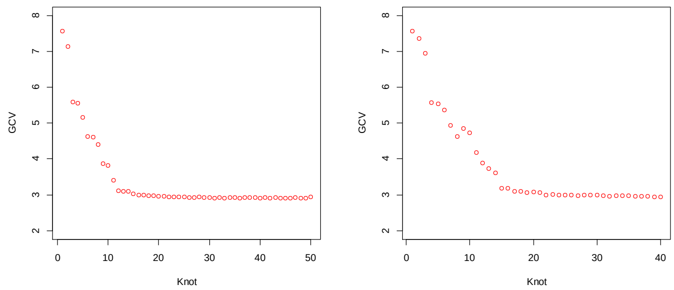
Figure 1 (a) showed a significant fluctuation change in data which form the diffraction intensity peaks. There are four peak of diffraction intensities, including around the angles 32, 44, 64 and 78 degrees with the highest peak around the 44 degree angle. The highest peak in Figure 1 (b) is higher than that of Figure 1 (a) and also the peak formed around the 32 and 64 degree angles in Figure 1 (b) is more gentle than in Figure 1 (a).

Table 1: Descriptive Statistics of Each Variable

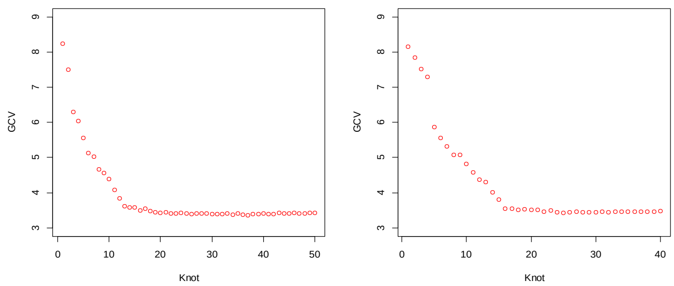
Variables	n	Min	Max	Mean	Std. Dev
X	3501	10	80	45	20.2159
Y_1	3501	0	32	2.2445	2.7567
Y_2	3501	0	38	2.5895	2.9022

Table 1 shows that the number of observations for each variable is 3501. The lowest *2θ* degree angle is 10 degrees and the highest is 80 degrees at intervals of 0.02 degrees. The diffraction intensity that occurs in $SrTiO_3$ materials is at least 0 times per second and the highest is 32 times per second. The average diffraction intensity that occurs is 2.2445 times per second with a standard deviation of 2.7567. For the diffraction intensity which occurs in $SrTiO_3+RuO_2$ material is at least 0 times per second and the highest at 38 times per second. The average diffraction intensity that occurs is 2.5895 times per second with a standard deviation of 2.9022.

The initial step in the formation of a spline trucanted regression model is to determine the number and location of the knots. Figure 2 shows the movement of GCV values for the number and location of knot points using the full-search algorithm. Figure 2 (a) and 2 (c) shows the movement of GCV values in $SrTiO_3$ and $SrTiO_3+RuO_2$ materials with linear models. Figure 2 (b) and 2 (d) shows the movement of GCV values in $SrTiO_3$ and $SrTiO_3+RuO_2$ materials with quadratic models.



(a) (b)



(c) (d)

Figure 2 : GCV Values for Each Knots

Based on Figure 2 (a) and 2 (c), it can be seen that the movement of GCV values has converged to the range of 2.8 to 3.6, so it can be concluded that for the GCV value with a greater number and location of knot points will be around that value. Thus it can be determined that the number and location of the knot points used for the formation of linear models is only up to 50 knots.

In Figure 2 (b) and 2 (d) it can be seen that the movement of GCV values has converged to the range of 2.9 to 3.7, so it can be concluded that for the GCV value with a larger number and location of knot points will be around that value. Thus it can be determined that the number and location of the knot points used for the formation of linear models is enough up to 40 knots.

The selection of the number and location of the optimal knot points is based on the number and location of the knot points which produce the smallest GCV value. The results of the calculation of GCV values on each material and model can be seen in Table 2.

Table 2: Minimum GCV values for each material and model

Material/Model	Number of Knot	GCV
$SrTiO_3$		
Linear	42	2.8991
Quadratic	40	2.9436
$SrTiO_3+RuO_2$		
Linear	37	3.3618
Quadratic	25	3.4303

Based on Table 2, the optimum number of knots in $SrTiO_3$ materials for linear models is 42 knots with a GCV value of 2.8991. For quadratic models, the optimum number of knots is 40 knots with a GCV value of 2.9436. Whereas in the $SrTiO_3+RuO_2$ material for the linear model, the optimal number of knots is 37 knots with a GCV value of 3.3618. For the quadratic model, the optimal number of knots is 25 knots with a GCV value of 3.4303.

After the number and location of the optimal knot points and the estimated of truncated spline regression coefficient for each model with the optimum number and point of knots formed, then the next step is to select the best model. The selection of the best model is based on the smallest AIC value and the largest R^2 value among the models. The results of the calculation of AIC and R^2 value can be seen in Table 3.

Table 3: The AIC and R^2 values of Models

Material/Model	AIC	R^2
$SrTiO_3$		
Linear	13403.33	0.6503
Quadratic	13469.34	0.6435
$SrTiO_3+RuO_2$		
Linear	13953.68	0.6302
Quadratic	14099.86	0.6132

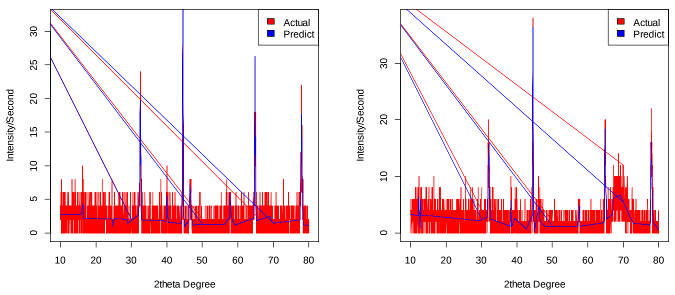
Table 3 shows that the best model for estimating the $SrTiO_3$ material regression curve is a linear model with 42 point knots. This is because the AIC value in the linear model is smaller than the quadratic model. In addition, the value of R^2 in the linear model is greater than the quadratic model. In the $SrTiO_3+RuO_2$ material, the best model for estimating the regression curve is a linear model with 37 point knot. This is because the AIC value in the linear model is smaller than the quadratic model. In addition, the value of R^2 in the linear model is greater than the quadratic model. The best model can also be seen from the accuracy of the model in predicting the data. Table 4 shows a comparison of actual data and truncated spline regression model predictive data for each material.

Table 4: Comparison of Actual and Predicted Data

Statistic	$SrTiO_3$		$SrTiO_3+RuO_2$	
	Actual	Predicted	Actual	Predicted
n	3501	3501	3501	3501
Minimum Intensity	0	0.9271	0	0.2209
Maximum Intensity	32	33.7252	38	36.3593
Position of X When Y is Maximum	44.48^0	44.48^0	44.56^0	44.56^0
Mean	2.2445	2.2445	2.5895	2.5895
Standard Deviation	2.7567	2.2304	2.9022	2.3112
MSE		2.6244		3.0799

Based on Table 4, the model used to prediction the regression curve were very good. It shown from the predicted data which not too different from the actual value. In addition, the MSE value generated by the two models were small.

The curves resulted by each model can be seen in Figure 3. Figure 3 (a) shows the curve of $SrTiO_3$ and Figure 3 (b) shows the curve of $SrTiO_3+RuO_2$.



(a) (b)

Figure 3 : Estimated Regresion Curve of (a) $SrTiO_3$ and (b) $SrTiO_3+RuO_2$

IV.CONCLUSION

The best model to estimate the curve of $SrTiO_3$ material was the linear truncated spline regression model with 42 knots, whereas the best model to estimate the curve of $SrTiO_3+RuO_2$ material was the linear truncated spline regression with 37 knots.

III. REFERENCES

[1]. Eubank, RL. 1999. Nonparametric Regression and Spline Smoothing 2nd ed. New York (US): Marcell Dekker, Inc.

[2]. Hardle, W. 1990. Applied Nonparametric Regression. New York (US): Cambridge University Press.

[3]. Ruppert, D. 2000. Selecting the Number of Knots for Penalized Splines. Journal of Computational and Graphical Statistics. doi: 10.1198/106186002853.

[4]. Spriti S, Eubank R, Smith PW, Young D. 2012. Knot Selection for Least Squares and Penalized Splines. Journal of Statistical Computation & Simulation. doi: 10.1080/00949655.2011.647317.

[5]. Suparti, Safitri D, Puspitasari I, Devi AR . 2013. Analisis Data Inflasi Di Indonesia Menggunakan Model Regresi Kernel. Prosiding Seminar Nasional Statistika [internet]; [waktu tidak diketahui]; Semarang, Indonesia. Semarang (ID): Universitas Diponegoro. pages 499 - 509 ;[download at March 22nd 2018]. Available at: http://eprints.undip.ac.id/40349/1/C09_Suparti.pdf

[6]. Tripena A. 2011. Analisis Regresi Spline Kuadratik. Prosiding Seminar Nasional Matematika dan Pendidikan Matematika [internet]; December 3rd 2011; Yogyakarta, Indonesia. Yogyakarta (ID): Universitas Negeri Yogyakarta;[download at March 22nd 2018]. Available at: http://eprints.uny.ac.id/7288/1/s-2.pdf

[7]. Widyawati N. 2012. Analisis Pengaruh Heating Rate Terhadap Tingkat Kristal dan Ukuran Butir Lapisan BZT yang Ditumbuhkan dengan Metode Sol Gel[skripsi]. Solo (ID): Universitas Sebelas Maret.