Scientia Africana, Vol. 23 (No. 4), October, 2024. Pp 123-132
© Faculty of Science, University of Port Harcourt, Printed in Nigeria *ISSN 1118 – 1931*

A HYBRID KERNEL: FAST-FOURIER TRANSFORM KERNEL MODEL

[1]Otaru, O.P., [2]Barguma, W.S and [3]Onoghojobi, B.
Department of Statistics
Federal University Lokoja, Nigeria
[1]otaru.paul@fulokoja.edu.ng; [2]simonbarguma@gmail.com and[3]benson.onoghojobi@fulokoja.edu.ng

Received: 23-08-2024
Accepted: 13-10-2024

https://dx.doi.org/10.4314/sa.v23i4.11

Journal Homepage: http://www.scientia-african.uniportjournal.info
Publisher: *Faculty of Science, University of Port Harcourt.*

ABSTRACT

In reality, most relationships between random variables are nonlinear, and forcing a linear fit can lead to inaccurate predictions and poor model performance. Linear models are unsuitable for most situations due to a lack of flexibility to capture complex patterns and issues with overfitting when data is complex and underfitting when the model is too simple to capture the underlying pattern in the data. Thus, this study compares the performance of the proposed Fast Fourier transform kernel regression with some nonlinear models using simulated and real-life data. The result shows that integrating kernel regression with the Fast Fourier transform proves to be an efficient model based on a minimum root mean square error value of 0.341 and a maximum coefficient of determination (R^2) value of 0.81 for simulated data, while real-life data had values of 0.007 and 0.94, respectively.

Keywords: Kernel Regression, Fast-Fourier Transform, Bandwidth, Algorithm, Nonlinear

INTRODUCTION

Regression is a statistical method used to investigate the effect of independent variables on dependent variables and the pattern of these relationships (Draper and Smith, 1998). In reality, many problems are associated with the relationship between the dependent and independent variables. Especially in economic situations where the forms of the relationship do not have a certain pattern, so they are not resolved, a regression approach is used with a nonlinear form, where the data pattern is assumed to be known (Hardle, 1990), or a nonparametric form, where the data pattern is not considered to follow any certain pattern (Wahyuni et al., 2020). Indeed, nonlinearity may well be expected in most human activity, given the complexity of the real world. Nonlinear relationships may be estimated by fitting nonlinear functions directly to the original data. However, the direct method usually involves highly complex computations if the relationships are nonlinear in the parameters (Koutsoyiannis, 1977). There are simple forms of nonlinear models, which can be linearized by appropriate transformations. Although some models cannot be linearized by transformation techniques, they are said to be intrinsically nonlinear in the parameters and variables (Nduka, 2010). Nonlinear models are more difficult to handle specification and estimation than linearized models. The difficulty in fitting the intrinsically nonlinear models varies with the complexity of the form of the model, the selected procedures, and the choice of initial guess values (which facilitate speedy convergence) based on iterative methods such as Gauss-Newton, steepest descent, and the Marquardt algorithm (Otaru and Iwundu, 2017). Adya and Collopy (1998), Chatfield (1995), and Tkacz (2001) found no clear evidence on whether nonlinear models might provide better forecasts than linear and random walk models when applied to financial time series data. In reality, not all data can be estimated with the parametric regression approach because there is no complete information about the shape of the regression curve; hence, a nonparametric regression approach can be used (Nur'eni et al., 2021).

Nonparametric estimators that are widely used are smoothing estimators such as kernel regression. Smoothing in nonlinear models is used to reduce noise, handle complex relationships, prevent over-fitting, enhance interpretability, and improve the continuity of the model. Also, it allows for a balance between model complexity and accuracy. Skabar (2008) used the kernel regression method to estimate density distributions of returns to generate out-of-sample forecasts. Niglio and Perna (2003) applied kernel regression with the corrected version of generalized cross-validation for optimal bandwidth selection to two climatic time series data collected from South Italy (Scafari) from January 1960 to December 2000. Again, one method that can be used in nonlinear regression is polynomial quadratic regression, while the method that can be used in nonparametric regression is kernel regression. The kernel regression is implemented using the smoothing technique, which is based on the kernel function used (Puspitasari, 2012; Vinod, 2017). Fast-Fourier transform Kernel regression is a technique that leverages the Fast-Fourier transform algorithm resulting in an efficient modeling in the frequency domain. This study aims to compare Fast-Fourier Transform (FFT) Kernel Regression and other nonlinear regression models based on the root mean sum of square RSME value and coefficient of determination (R^2).

MATERIALS AND METHOD

Source of Data

A noisy signal from a sensor (i.e., vibration from a machine) was simulated. The signal contains high-frequency noise that needs to be filtered out to obtain the underlying trends or patterns. Also, the study adopted real-life data from a chlorine experiment by Draper & Smith (1998). These data were analyzed using a statistical package called R-Language.

Kernel Regression

Kernel regression is a non-parametric technique that estimates the conditional expectation of a random variable. The main objective is to estimate the functional non-linear relation between a pair of random variables X and Y. The conditional expectation of a variable Y relative to a variable X can be expressed as:

$$E(Y/X) = m(x) \tag{1}$$

where $m(x)$ is an unknown function to be estimated. Mathematically, for any value x, the smoothing estimator for m(x) can be expressed as follows

$$\hat{m}(x) = \frac{1}{n}\sum_{i=1}^{n}\omega_{hi}(x)Y_i \tag{2}$$

where ω_{hi} can be defined as a weighted function

$$\omega_{hi} = \frac{K_h(x - X_i)}{\hat{f}_h(x)} \tag{3}$$

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i-1}^{n}K_h(x - X_i) \tag{4}$$

then the density estimator

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x - X_i}{h}\right) \tag{5}$$

The general form of K (x) kernel in terms of bandwidth usage

$$K_h(x) = \frac{1}{h}K\left(\frac{x}{h}\right) \tag{6}$$

Scientia Africana, Vol. 23 (No. 4), October, 2024. Pp 123-132
© Faculty of Science, University of Port Harcourt, Printed in Nigeria *ISSN 1118 – 1931*

where k is a kernel function, n is the number of observation or data points and h is the bandwidth or smoothing parameter.

Characteristics of a Kernel Function

i. $\int_{-\infty}^{\infty} K(x)\,dx = 1$

ii. $\int_{-\infty}^{\infty} x\,K(x)\,dx = 0$

iii. $\int_{-\infty}^{\infty} x^2 K(x)\,dx = \mu_2(K) \neq 0$

iv. $\int_{-\infty}^{\infty} [K(x)]^2\,dx = \int_{-\infty}^{\infty} K^2(x)\,dx = \|K\|_2^2$

By substituting (3) into (2), the Nadaraya-Watson kernel estimator from m(x) is obtained as follows (Nadaraya, 1964).

$$\hat{m}_h^{NW} = \frac{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) Y_i}{\sum_{j=1}^{n} K\left(\frac{x - X}{h}\right)} \quad (7)$$

The choice of kernel function is the Gaussian kernel namely

$$K_G(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) \quad (8)$$

Optimum Bandwidth Selection

The optimal bandwidth is very important, as small and large bandwidth values result in rough estimation curves and very smooth estimation curves, respectively. The study adopted the plug-in method based on an extension of the mean integrated square error (MISE) for kernel smoothing given as (Wand, 1994; Scott, 2015).

$$MSE\left|\hat{f}_h(x)\right| = \frac{1}{nh}\|K\|_2^2 + \frac{h^4}{4}\mu_2(K)^2\|f''\|_2^2 + 0((nh)^{-1}) + 0(h^4) \quad (9)$$

Then the asymptotic MISE (A-MISE) is obtained by ignoring the use of $0((nh)^{-1}) + 0(h^4)$, so

$$A - MISE = \frac{1}{nh}\|K\|_2^2 + \frac{h^4}{4}\mu_2(K)^2\|f''\|_2^2 \quad (10)$$

So that the optimum bandwidth (h_{opt}) is

$$h_{opt} = \left(\frac{\|K\|_2^2}{\|f''\|_2^2(\mu_2(K))^2 n}\right)^{1/5} \quad (11)$$

where $\mu_2(K)$ is the second moment of the kernel, $\|f''\|_2^2$ is the second derivative of the kernel when n tends to infinity and $\|K\|_2^2$ is the rugosity of the kernel. Therefore by Silverman's rule (Silverman, 1986)

$$h_{opt} = \frac{1.06}{n^{1/5}} \delta \tag{12}$$

where δ is the standard deviation and n is the number of observations

Fast-Fourier Transform

Fast-Fourier transform (FFT) computes the discrete Fourier transform and its inverse. This is used to convert a digital signal (x) with length (N) from the time domain into a signal in the frequency domain (Ali et al., 2024).

The Fourier transform $\tilde{f}$ of an integrable function $f : R^g \to C$ with $s \in R^g$ is given by

$$\tilde{f}(s) = \int_{R^g} f(x) \exp\left(-2\pi i < x, s >\right) \partial x \tag{13}$$

under suitable conditions, the inverse transform from $\tilde{f}$ to f with $t \in R^g$ is given by

$$f(t) = \int_{R^g} \tilde{f}(x) \exp\left(-2\pi i < s, x >\right) \partial s \tag{14}$$

Properties of Fourier Transform

Assume $f(x)$, $g(x)$ and $h(x)$ are integrable functions:

Linearity: For $a, b \in C$ if $f(x) = ah(x) + bg(x)$ then $\tilde{f}(s) = a\tilde{h}(s) + b\tilde{g}(s)$ (15)

Translation: For $x_0 \in R^n, if\ f(x) = h(x - x_0)$ then $\tilde{f}(x) = \exp(-2\pi i < x_0, s >)\tilde{h}(s)$ (16)

Modulation: For $s_0 \in R^n, if\ f(x) = \exp(-2\pi i < x, s >)h(s)$ then $\tilde{f}(s) = \tilde{h}(s - s_0)$ (17)

Sealing: For $a \neq 0\ if\ f(x) = h(ax)$, then $\tilde{f}(s) = \frac{1}{\left|\prod_{i=1}^{g} a_i\right|} \tilde{h}\left(a^{-I} s\right)$ (18)

where $a \in R^g, a^{-I} = \left(a_1^{-I}, \cdots, a_g^{-I}\right)$

Convolution Theorem

An important feature of the Fourier transform is convolution. Suppose two function f and g is given, then the convolution can be defined as (Ali et al., 2024):

$$\left(f * g\right)(z) = \int_{R^g} f(x) g(z - x) \partial x$$ then

$$\tilde{f} * \tilde{g} = \tilde{f} \,.\, \tilde{g} \tag{19}$$

Scientia Africana, Vol. 23 (No. 4), October, 2024. Pp 123-132
© Faculty of Science, University of Port Harcourt, Printed in Nigeria *ISSN 1118 – 1931*

To speed up the computation of kernel density estimators using the FFT, let $x_1, \cdots, x_n$ be a random sample drawn from an unknown distribution with density f. Given that the kernel density estimator with bandwidth h is

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) \tag{20}$$

then $(\tilde{f} * \tilde{g})_n = \tilde{f}_n \cdot \tilde{g}_n$ (21)

Fourier transform can speed up convolutions by taking advantage of the property given as

$$x(n) * y(n) = \sum_{k=-\infty}^{\infty} x(k)y(n-k) \overset{DFFT}{\leftrightarrow} X(e^{jw})Y(e^{jw}) \tag{22}$$

where y is dependent and x is independent variable

The equation states that the convolution of two signals is equivalent to the multiplication of their fourier transforms. Fast fourier transform (FFT) is used to improve the computation of a kernel regression. A popular estimator for the regression function $E(X/Y) = m\,(x)$ given observed points $(x_1, y_1), \cdots, (x_n, y_n)$ is the Nadaraya-Watson estimator with Kernel function $K_h\,()$ and bandwidth h.

$$\hat{m}_h\,(x) = \frac{\sum_{i=1}^{n} K_h\left(\frac{x - X_i}{h}\right) Y_i}{\sum_{j=1}^{n} K_h\left(\frac{x - X_j}{h}\right)} \tag{23}$$

The FFT requires a grid of with $n' = 2^k, k \in N$ design points. A meaningful choice in terms of asymptotic is $K_h = \frac{\log(n)}{\log(2)}$. Then $\underset{n\to\infty}{Lim}\frac{2^k}{n} = 1$ which means that $n\,and\,n'$ are asympototically equivalent. In a first step we have to interpolate $y = (y_1, \cdots, y_n),\ x = (x_1, \cdots, y_n)$ to fit to an equidistant grid with n' points with $y' = y_a + (y_b - y_a)\,\frac{x' - x_a}{x_b - x_a}$ at the point x'

Since we use an interpolation, using an equidistant grid $x' = \left(x_1', \cdots, x_n'\right)$ become

$$\hat{m}_h'\,(x) = \frac{\max(x') - \min(x')}{n'} \sum_{i=1}^{n'} K_h(x - X_i) \tag{24}$$

Let the discrete Fourier transform of $\hat{m}$ be denoted by $\tilde{m}$. The Fourier transform of equation (24) for $s \in \left(x_1', \cdots, x_{n'}'\right)$ using convolution and translation is then given by

$$\hat{m}_h'\,(x) = \frac{\max(x') - \min(x')}{n'} \sum_{i=1}^{n'} K_h(s) y_i' e^{isx_i'} = \frac{\max(x') - \min(x')}{n'} \tag{25}$$

where $\tilde{y}(s) = \frac{1}{n'} \sum_{i=1}^{n'} y_i{}' e^{isx_i{}'}$ is the fast fourier transform of the interpolated data. Again, $\hat{m}'$ can then be obtained by applying the inverse discrete fast fourier transform to equation (25). All these steps involve only $O(n' \log(n'))$ operations instead of $O(n'n)$ operations when computing equation (24) directly. Sorting which may require prior to the linear interpolation has a computational complexity of $O(n \log(n))$ which has the same order of magnitude than estimating the FFT based estimator itself and thus does not impact the computational complexity.

Proposed Fast-Fourier Transform (FFT) Kernel Regression

The FFT-based kernel regression algorithm:

1. Data is obtained (independent (x) and dependent (y) variables)
2. Define a FFT-based kernel regression function that takes independent variable (x), corresponding dependent variable (y)
3. Inside the function, compute the FFT of the data and a Gaussian kernel
4. Perform point-wise multiplication of FFTs in the frequency domain (multiply the frequency-domain representations of the data and the kernel function)
5. Perform an inverse FFT of the multiplied spectrum (to obtain the result in the time domain or smoothed data)
6. Choose an appropriate kernel bandwidth parameter (optimum bandwidth)
7. Apply the FFT-based kernel regression function to smooth the data
8. Obtain the root mean square error (RMSE), coefficient of determination (R^2) and plot smoothed data using FFT-based kernel regression against original data.

RESULTS AND DISCUSSION

Data was subjected to a normality test to investigate the pattern. The result reveals that noise signal and chlorine are not normally distributed, thus a nonlinear regression model is appropriate for both data (see Table 1). Evaluating the models, the results of FFT-based kernel regression reveal significant ($p<0.05$) least RMSE and higher R^2 values with a bandwidth of 0.3, followed by a kernel regression model with a bandwidth of 0.5, a polynomial model (second order), respectively, and an insignificant ($p>0.05$) linear model (see Figs. 1 & 2). Prediction based on the models revealed close prediction values for FFT-based kernel and kernel regression (see Tables 3 & 4). This proposed method shows the performance of regressive techniques such as kernel when merged with a technique that identifies patterns or periodicities in data that might not be obvious in the time domain. Similar models are random compact Gaussian (RCG) kernel (Xiao-jian et al., 2012), Bayesian kernel machine regression (BKMR) (Bobb et al., 2018), extreme learning machine (ELM) (Huang et al., 2006), Choquet kernel (Ali, 2021), hybrid approach based on kernel regression (Lomowski& Hummel, 2020), quantum kernel methods (Annie et al., 2023), kernel inducing points (KIP) (Timothy et al., 2020), and kernel-based dynamic programming (DP) (Ali et al., 2024).

Table 1: Normality Test

	Kolmogorov-Smirnov		**Shapiro-Wilk**	
Test	**Noise Signal**	**Chlorine**	**Noise Signal**	**Chlorine**
Statistic	0.201	0.279	0.892	0.856
Degree of Freedom	100	18	100	18
p-value	0.000	0.001	0.000	0.011

Scientia Africana, Vol. 23 (No. 4), October, 2024. Pp 123-132
© Faculty of Science, University of Port Harcourt, Printed in Nigeria *ISSN 1118 – 1931*

Table 2: Model Evaluation

Regression	Noise Signal RMSE	R-Square	Chlorine RMSE	R-Square
Linear	0.8012154	0.00004698	0.01509405	0.7382513
Polynomial	0.7409249	0.1448754	0.01115869	0.8569463
Kernel	0.4402131	0.6981391	0.00736639	0.9376579
FFT-Kernel	0.3411079	0.8187557	0.00736638	0.9376579

Table 3: Prediction for Noise Signal

Sensor	Linear	Polynomial	Kernel	FFT-Kernel
10.1	0.1339884	0.8361646	-0.1938564	-0.2359301
10.3	0.1336117	0.9192069	-0.2480801	-0.1803068
10.5	0.1332349	1.005458	-0.2824959	-0.1714699
10.7	0.1328582	1.094917	-0.3022911	-0.1702687
10.9	0.1324815	1.187584	-0.3118485	-0.1701089

Table 4: Prediction for Chlorine

Weeks	Linear	Polynomial	Kernel	FFT-Kernel
44	0.3665157	0.4033098	0.39	0.39
46	0.3611581	0.4078416	0.39	0.39
48	0.3558006	0.4133087	0.39	0.39
50	0.3504431	0.419711	0.39	0.39
52	0.3450855	0.4270485	0.39	0.39

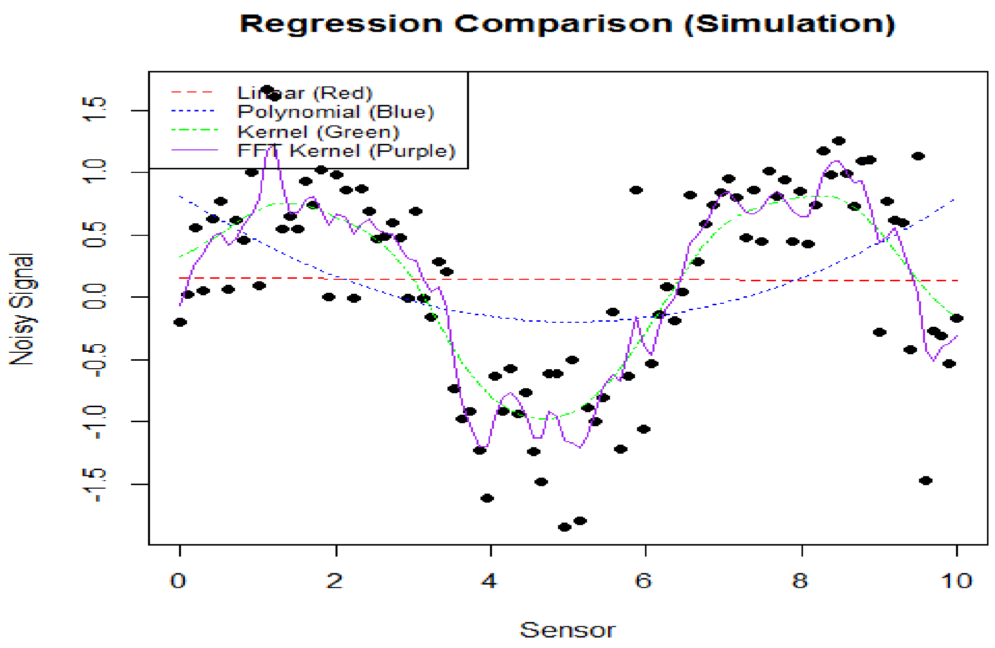

Fig. 1: Model Comparison Using Simulated Data

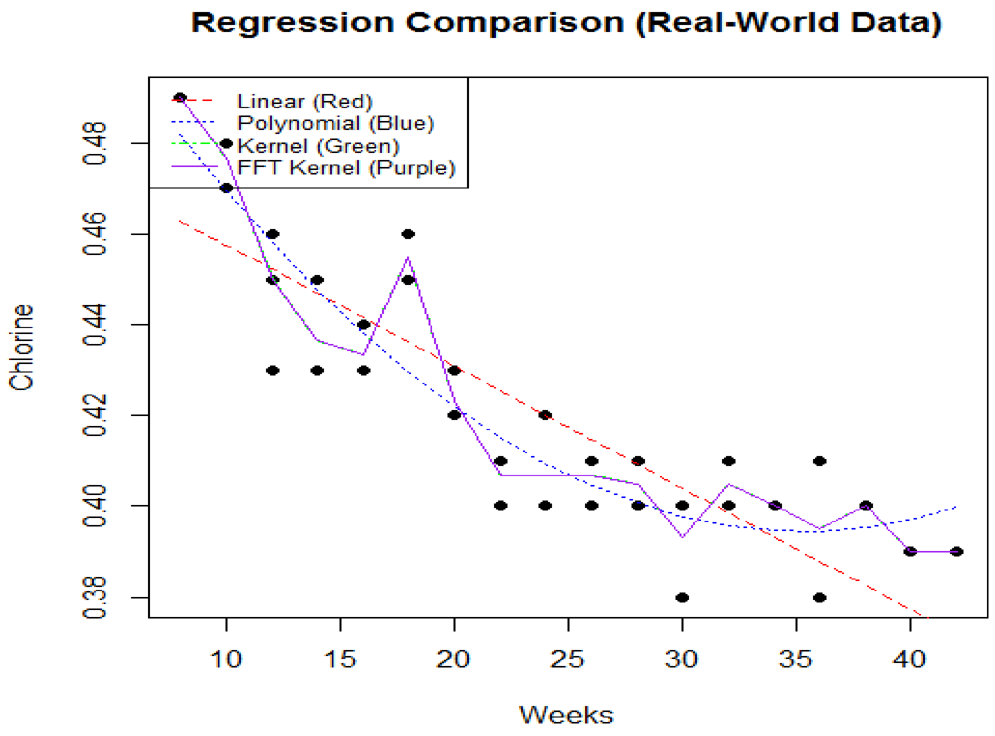

Fig. 2: Model Comparison Using Real-Life Data

CONCLUSION

This study examined the performance of the Fast-Fourier Transform (FFT) kernel regression model over other nonlinear regression models for simulated and real-life data. Fast-Fourier Transform (FFT) kernel regression model revealed a strong positive coefficient of determination measure between chlorine and age (weeks), followed by the kernel regression model. By using the Fast-Fourier Transform (FFT), the computational complexity of convolution operation (which is typically O (n^2) in the time domain) can be reduced to O (n/log n), making it much more efficient, especially for large datasets where the reduction in computation time is significant. Also, it assumes a specific parametric form for the underlying data distribution, making it highly flexible. However, FFT-based kernel regression may be efficient for certain kernels and datasets based on an optimum bandwidth; it might not be suitable for all datasets. Nevertheless, the Fast-Fourier Transform (FFT) kernel regression model in comparison with kernel regression and the polynomial (second order) regression model based on the root mean square of the square (RSME) and coefficient of determination (R^2) proves to be more efficient and flexible.

REFERENCES

Adya, M. and Collopy, F. (1998). How effective are neural networks at forecasting andprediction? A review and evaluation; *J. Forecasting*, 17: 481–495.

Ali F. T. (2021). The choquet kernel on the use of regression problem. Information sciences, 556: 256-272.

Ali, F., Raffaele, L., Massimo, T, and Silvio, B. (2024). A kernel based approximate dynamicprogrammingapproach: Theory and application. Automatica 162, 111517.

BobbJ.F, Henn B.C., Valen L., CoullB.A (2018). Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel

Scientia Africana, Vol. 23 (No. 4), October, 2024. Pp 123-132
© Faculty of Science, University of Port Harcourt, Printed in Nigeria *ISSN 1118 – 1931*

machine regression. Environ Health 17 (1), 67.

Chatfield C (1995). Positive or negative? *Int. J. Forecasting*, 11: 501–502.

Draper .N. R., and Smith .H. (1998). Applied regression analysis (*New York: John Wiley & Sons*)

Hardle .W.,andVieu .P. (1991). Kernel regression smoothing of time series, *J. Time Series Anal.,* 13: 209 223.

Huang .G.,Zhiu .Q., and Kheong .S. (2006). Extreme learning machine: Theory and applications. Neurocomputing 70 (1-3) 489-501.

Lomowski, R. and Hummel, S. (2020). A method to estimate the remaining useful life of a filter using a hybrid approach based on kernel regression and simple statistics. Proceedings of the European conference of the Prognostics and Health Management Society (PHME), 5 (1). Doi:10.36001/phme. 2020v5i1.1316.

Nadaraya .W. (1964). On estimating regression. *Probability Theory and its Applications* 10: 186-190.

Niglio .M., and Perna .C. (2003). Kernel smoothing for the analysis of climatic data, *Quadernidistatistica*, 5.

Nur'Eni, Fajri .M., and Astuti .S. (2021). Comparison of Kernel regression model with a polynomialregression model on financial data. *Journal of Physics:* Conference Series 1763,012017, IOPdoi:10.1088/1742-6596/1763/1/ 012017

Puspitasari .I.,Suparti .S., and Wilandari .Y. (2012). Jurnal Gaussian 1(1) 93-102

Scott, D. N (2015). Multivariate density estimation: theory, practice, and visualization. John Wiley & Son,

Silverman, B. W (1986). Density estimation. Chapman & Hall, London.

Skabar .A. (2008). A kernel-based technique for direction-of-change financial time series forecasting. J.Fin. Time Series Forecasting, 14(4): 275-284.

Timothy .N.,Zhourong .C., and Jaehoon .L. (2020). Dataset-meta-learning from kernel ridge regression. arXiv Preprint arXiv:2011.00050

Tkacz .G. (2001). Neural network forecasting of Canadian GDP growth, Int. J. Forecasting, 17:57–69.

Vinod .H. D. (2017). Generalized correlation and kernel causality with applications in development economics. Communications in Statistics—Simulation and Computation 46, 4513–34.

Wahyuni .S. A, Ratnawati .R.,Indriyani .I., and Fajri .M. (2020). Natural Science: Journal of Science and Technology 9(2) 34-39.

Wand, M.P and Jones, M. C (1994). Kernel smoothing, First Edition, Chapman & Hall, London

Xiaojian .D.,Jian .L.,Fan Y., and Jie .C. (2021). Random compact Gaussian kernel: application to ELM classification and regression. Knowledge based systems, 217, 106848.