BIU Journal of Basic and Applied Sciences 6(1): 65 – 77, 2021.

ISSN: 2563-6424

A NOVEL JOINT ADAPTIVE POLYNOMIAL REGRESSION AND WAVELET TRANSFORM METHOD FOR CORONAVIRUS EPIDEMIC TREND RATE ESTIMATION

***OJUH, O. D.[1] AND ISABONA, J.[2]**
[1]Department of Physical Sciences, Benson Idahosa University, Benin City, Edo State
[2]Department of Physics, Federal University Lokoja, PMB 1154, Kogi Sate, Nigeria
*Corresponding author: dojuh@biu.edu.ng

ABSTRACT
Predictive trend pattern studies using robust adaptive regression methods remained an important step towards assessing and revealing the prognostic connections between different variables in real-time observations. In this work, a polynomial regression technique scaled with wavelet transform is proposed to model and investigate coronavirus epidemic trend rate in Nigeria. Various statistical tests employed to ascertain if the proposed regression modeling and estimation approach is satisfactory are also contained in the wok. The resultant statistical outcome shows that the fifth degree polynomial model with joint discrete wavelet transform scheme is the most appropriate regression method to model and estimate the Coronavirus epidemic data trend rate. The results obtained using the ordinary polynomial regression models are also provided as performance benchmarks.

KEYWORDS: Adaptive regression, Polynomial, Wavelet transform, Coronavirus, Trend estimation

INTRODUCTION
Trend pattern studies with regression models have been used largely in different fields of engineering, physical sciences, social sciences and medical sciences to investigate both long term and short connections between or among variables in datasets: weather variables, health mortality variables, myocardial infarction variables, environmental exposure variables and among others (Jimenez *et al*., Basu, 2009; Armstrong, 2006; Bell *et al*., 2004; Schwartz *et al*., 1996 and Bhaskaran *et al*., 2010).

In physical and mathematical sciences, regression is one of the most extensively employed approaches to determine correlation pattern between one or many independent variables or dependent variables, conduct predictive analysis of future events and provide a robust description of data structure pattern. There are several regression models in literature. Among the common ones are linear regression, logistic regression, stepwise regression, Lasso regression, ridge regression, polynomial regression and elastic regression, quantile regression, partial least square regression, cox regression, poison regression, principle component analysis regression, tobit regression, ordinal regression, Quasi poison regression, etc. These various regression model have been engaged by different researchers for diverse investigative studies. For example, in Bhaskaran *et al*. (2010), the influence of temperature on myocardial infarction in Wales and England using time series linear regression is presented. In Isabona and Enagbonma, (2014), the authors showed how the least absolute regression model can be used to conduct predictive analytic study and reduce signal prediction error in wireless networks. In Isabona and Enagbonma (2014); Isabona *et al.* (2013) and Isabona and Isaiah, (2015), the researchers employed diverse techniques to reveal how ordinary least square regression method can be explored to model and perform predictive analysis of signal attenuation loss data in telecommunication networks.

Particularly, the authors in Ogundokun *et al*. (2020) explored ordinary linear regression conduct predictive trend analysis study of confirmed cases corona virus in Nigeria. By means of multiple linear regression, the authors in Rath *et al.* (2020) also conducted a similar predictive analysis of corona infection, using Odisha and India as case study. The application of different prediction methods such as forward selection, subset selection, Lasso regression and Ridge regression models in estimating COVID-19 diseases cases is contained in (Qin *et al.,* 2020).

The main problem with ordinary least square regression method as used largely in the above studies is that it often over-simplifies real-world problems by assuming a linear correlation between the data input and output variables. This often leads to large predictive errors between data input and resultant output (Isabona, 2020).

This paper aims to propose and apply an adaptive polynomial regression modelling technique jointly scaled with mathematical physics-based discrete wavelet transform for better modelling and estimation of coronavirus epidemic trend rate in Nigeria. Various statistical tests computed in Matlab software environment are employed to ascertain if the proposed hybrid approach is satisfactory, are also provided in the wok.

THEORETICAL FRAMEWORK

Polynomial regression model is a special type of statistical regression models that provides means to adaptively determine the bond between quantitative input variables (independent variables) and the output variables (dependent variable). A polynomial modeling dependent variable x can be of order two (Quadratic), three (Cubic), four (Quartic), or five order (Quintic) polynomials as shown in equations (1) to (4).

$$\varepsilon(y) = c_o + c_1 x + c_2 x^2 + \varepsilon \quad (1)$$

$$\varepsilon(y) = c_o + c_1 x + c_2 x^2 + c_3 x^3 + \varepsilon \quad (2)$$

$$\varepsilon(y) = c_o + c_1 x + c_2 x^2 + c_3 x^3 + c4 x^4 + \varepsilon \quad (3)$$

$$\varepsilon(y) = c_o + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 + c_5 x^5 + \varepsilon \quad (4)$$

where ε is an unobserved random error term, and $c_0, c_1, \ldots$ are unknown parameters. Generally, higher-order polynomials can be expressed as:

$$\varepsilon(y) = c_o + c_1 x_i + c_2 x_i^{2=} + c_3 x_i^3 + \ldots; c_k x_i^k + \varepsilon_{i_(} (i = 1,2,3,\ldots,n) \quad (5)$$

where i and x_i the observation index and the explanatory dependent variable. The expression in equation (5) is also called a k[th] degree order polynomial model. In this work, k[th] degree order polynomial is termed adaptive polynomial (AP) because it possess the ability to adaptively model non-linear datasets as the degree polynomial order increases.

In matrix form, equation (5) can expressed as:

$$Y = Xc + \varepsilon \quad (6)$$

where,

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 x_{1,1} \ldots x_{j,1} \\ 1 x_{1,2} \ldots x_{j,2} \\ \vdots \quad \vdots \quad \vdots \\ 1 x_{1,n} \ldots x_{j,n} \end{bmatrix}, c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} and\ \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Thus,

$$Y = Xc + \varepsilon = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 x_{1,1} \ldots x_{j,1} \\ 1 x_{1,2} \ldots x_{j,2} \\ \vdots \quad \vdots \quad \vdots \\ 1 x_{1,n} \ldots x_{j,n} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (7)$$

In terms of the error estimation parameter, Equation (7) can be written as:

$$\varepsilon = Y - Xc \quad (8)$$

By means of least square estimation technique, the unknown polynomial regression coefficients and their variance can be obtained by:

$$\varepsilon^T \varepsilon = (Y - Xc)^T (Y - Xc) \tag{9}$$

where $(\;)^T$ indicates matrix transposition.

By taking the derivative of the error term with respect to c and setting to zero, we have:

$$c = \left(X^T X\right)^{-1} X^T Y \tag{10}$$

$$Var(c) = \frac{\sigma^2}{\left(X^T X\right)^{-1}} \tag{11}$$

Physics of Wavelet Transform

A wavelet is a distinctive wave-like oscillation tool with local minimum, local maximum and amplitude that originates at zero point, rises, and then drops back to zero. As a mathematical physics tool, wavelets can be explored to detect, and extract, information of different data types. It can also be largely used to preprocess, decompose, compress and donoise data different sizes. Sets of discrete wavelets are usually required to analyze data completely.

Consider a typical complex-valued function, ψ which satisfies the conditions given by (Isabona, 2020):

$$\int_{-\infty}^{\infty} \psi|(t)|^2 dt \tag{12}$$

$$d_\psi = 2\pi \int_{-\infty}^{\infty} \frac{\Psi\left|(\psi)^2\right|}{|\omega|} d\omega \tag{13}$$

In equations (12) and (13), ψ stands for the Fourier transform of ψ and $\psi(t)$ indicates the wavelet function.

If ψ satisfies the above expressed conditions, then the wavelet transform of the field signal strength, $q(t)$ in terms $\psi(t)$ can be articulated as:

$$Q(n,m) = \frac{1}{\sqrt{m}} \int_{-\infty}^{\infty} \psi' \left(\frac{t-n}{m}\right) q(t) dt \tag{14}$$

where, ψ^{t} indicates the complex conjugate of ψ, *a* and *b* represent the scale and position parameters respectively. $q(t)$ indicates the field signal strength to be transformed

By expressing $\psi_{m,n}(t)$ as:

$$\psi(m,n) = \frac{1}{\sqrt{m}}\psi\left(\frac{t-n}{m}\right) \quad (15)$$

Then equation (15) can be written as scalar product or an inner product of the field signal, $q(t)$ such that function $\psi_{m,n}(t)$ gives:

$$q(n,m) = \frac{1}{\sqrt{m}}\int_{-\infty}^{\infty}\psi^{t}{}_{m,n}(t)q(t)dt \quad (16)$$

where $m \in \mathbb{R}^{+}$, $n \in \mathbb{R}$, with $m \neq 0$, and $\psi(\cdot)$ satisfies the condition for admissibility.

MATERIALS AND METHODS

Data Collection

We use Worldometer official website (https://www.worldometers.info/coronavirus/country/ nigeria/) to collect of Corona virus epidemic data in Nigeria. The website contain among other things, daily and monthly cases of coronavirus infections and the total number of deaths records from the epidemic of all the countries in the world. In this work, we only extracted data on daily cases of coronavirus infections in Nigeria, starting from 10 march, 2020 to August 12, 2020.

Data Preprocessing

Usually, every datasets always contains either noisy components or outliers, which must be removed (preprocessed) to enable their effective analysis. In this work, we explore wavelet transform to preprocess the acquired daily coronavirus pandemic data. The features of the wavelet transform is provided as follows: In the wavelet transform process, the Wden function tool in Matlab 2018a software has been engage and it has three implementation phases as highlighted below:

- Decomposition phase— Here, symlet wavelet of level 8 (i.e sym8) is chosen for the decomposition
- Thresholding phase — At this phase, SURE (Stein's Unbiased Risk Estimate) thresholding with soft thresholding rule is selected
- Reconstruction phase — the wavelet reconstruction computation is built on the original approximation coefficients.

Polynomial Regression Models and Procedure for order Selection

As revealed in equations (1) to (5), the polynomial regression models are of different orders. In this work, we employed the ‘forward selection method. As the method implies, we successively select and fit the polynomial models in an increasing order starting with order two (quadratic) and test the regression coefficients at steps and in forward manner to higher orders until a non-significant t-test value is attained.

Proposed Model Evaluation

To examine the estimation accuracy of proposed polynomial modelling approach, three statistical indexes, which are normalized root mean squared error *NRMSE*, adjusted coefficient of determination R^2 and none adjusted (i.e. the normal) coefficient of determination R^{*2}. The closer both R^2 and R^{*2} values are, the better the correlation accuracy. On the other hand, lower NRMSE values, the healthier the modelling accuracy.

RESULTS AND DISCUSSION

All the regression modelling based estimations, statistical computational analysis and graphics were accomplished using Matlab 2018a software. Displayed chart in figure 1 is the daily COVID-19 Infection rate data obtained from March 19 to August 12, 2020. The chart first shows an increasing trend of COVID-19 daily Infection rate till June 18, 2020 where it attains its pick at 775, and then before showing a decreasing trend. Shown in figures 2 to 4 are COVID-19 Infection rate trend estimation made using ordinary adaptive polynomial (AP) and the proposed combined wavelet transform and adaptive polynomial (WT-AP) for second degree order (quadratic) fifth degree order(quantic). The results indicate the estimation accuracies improves as the degree order of the polynomial increases. The respective residual errors plots and Normal distribution fitting on COVID-19 Infection rate with AP and WT-AP model estimations are provided in figures 5 to 8. Residual plots are provided to reveal how the proposed WT-AP model satisfactorily captures estimate the COVID-19 trend rate in comparison with the ordinary AP model. Based on the summarized statistical results in Tab. 1 to 3, the most appropriate regression model to estimate COVID-19 epidemic data trend rate is the fifth degree polynomial regression model combined with wavelet transform scheme. Compared to higher RMSE result of 77.94dB value with the ordinary polynomial regression, the proposed WT-AP model yielded a lower RMSE value of 55.67dB. Similar higher RMSE values were recorded in COVID-19 predicted work by Qin *et al.,* (2020), using similar methods such as Ridge regression and Lasso regression.

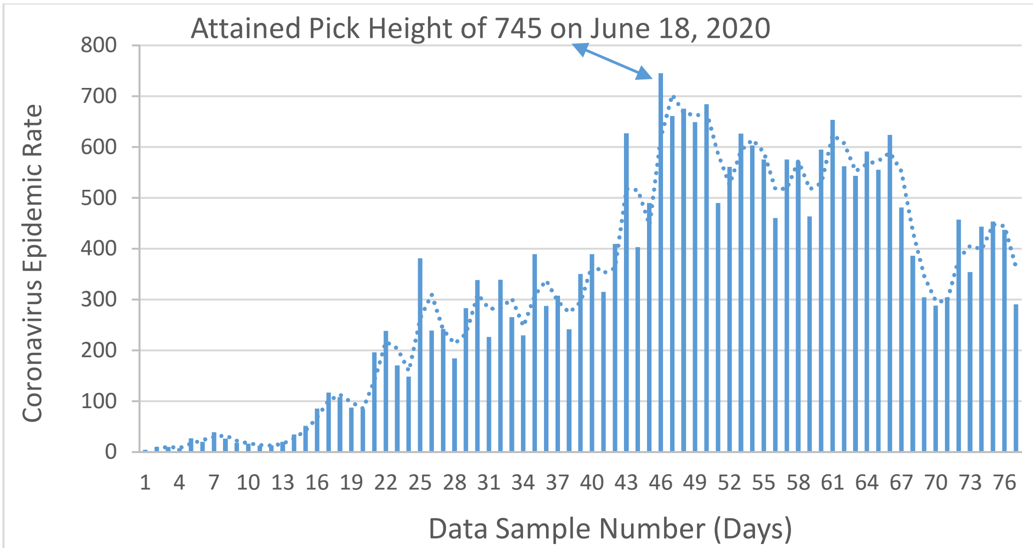

Fig. 1: COVID-19 Infection rate data obtained from March 19 to August 12, 2020

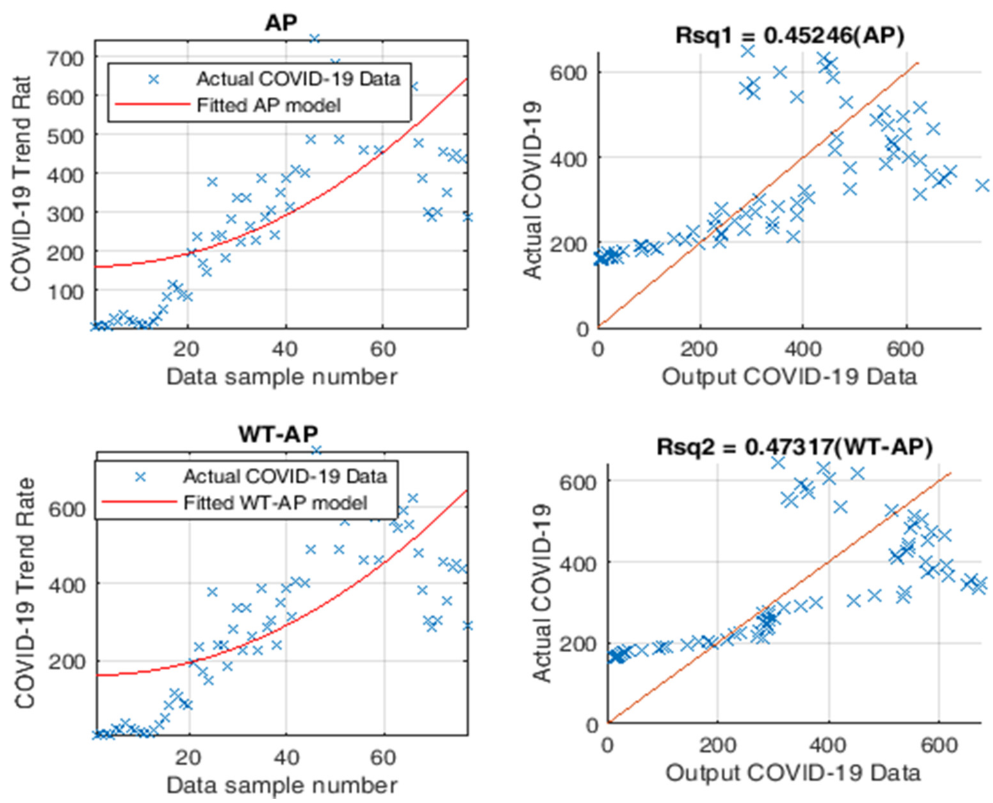

Fig. 2: COVID-19 Infection rate trend estimation with AP and WT-AP and their Correlation for Quadratic Polynomial order

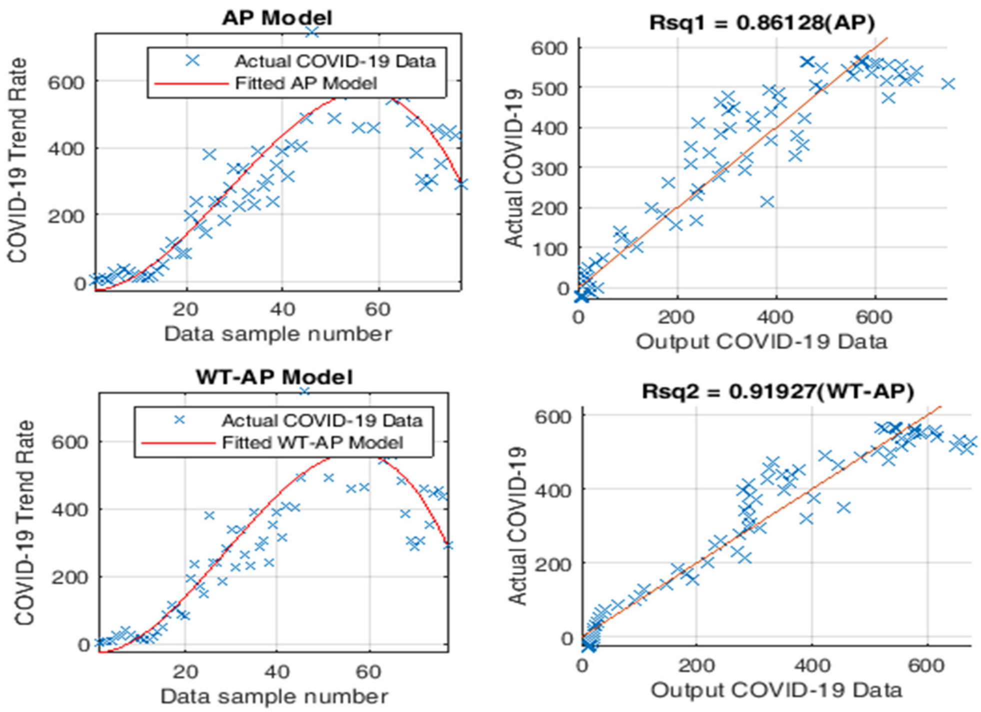

Fig. 3: COVID-19 Infection rate trend estimation with AP and WT-AP and their Correlation for Cubic Polynomial order

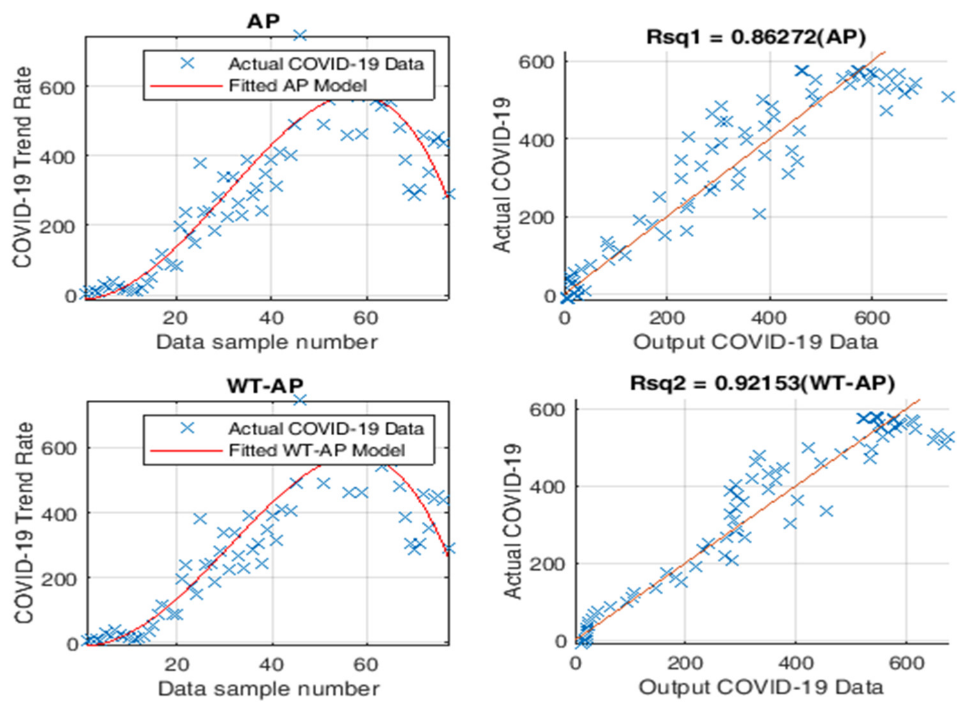

Fig. 4: COVID-19 Infection rate trend estimation with AP and WT-AP and their Correlation for Quartic Polynomial order

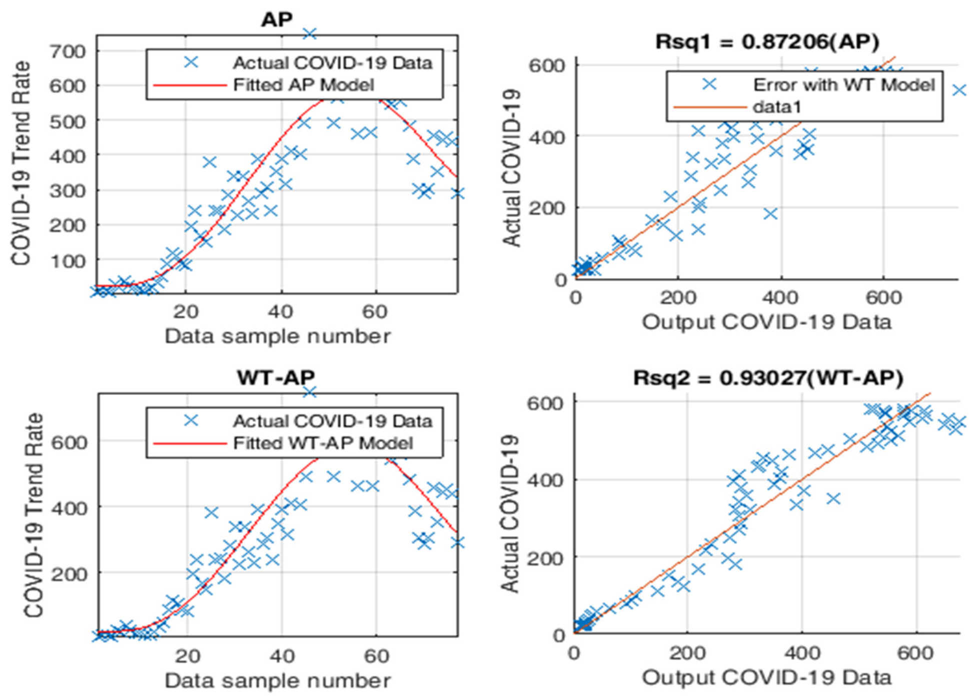

Fig. 5: COVID-19 Infection rate trend estimation with AP and WT-AP and their Correlation for Quintic Polynomial order

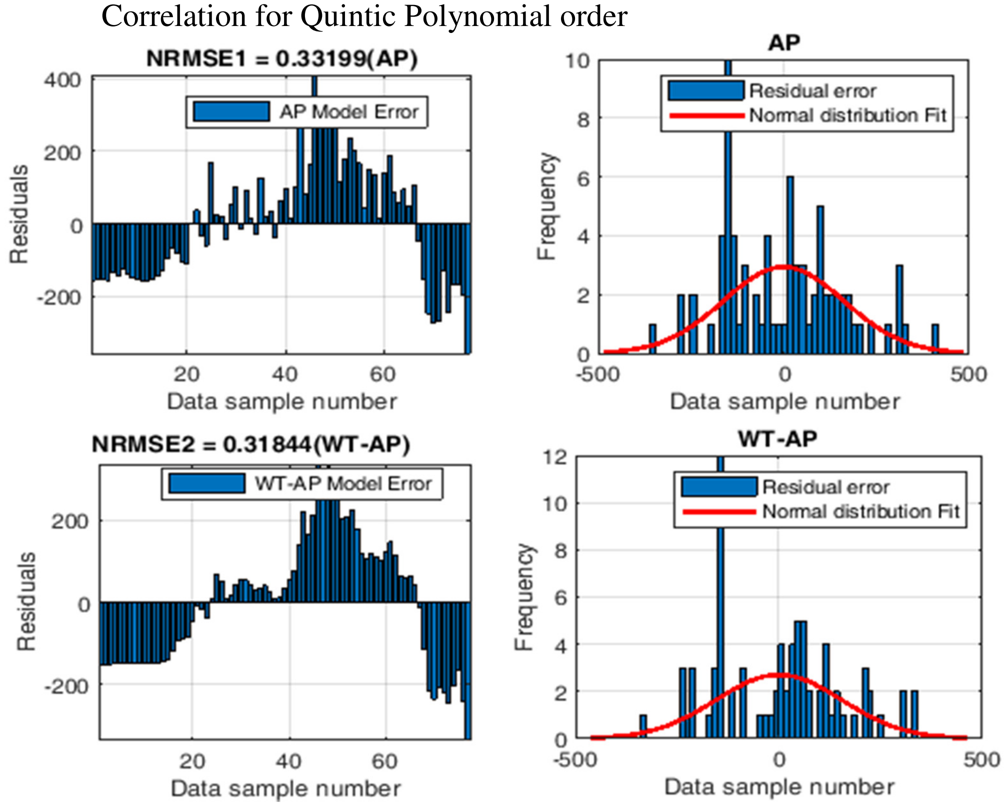

Fig. 6: Residual error plots of COVID-19 Infection rate trend estimation with AP and WT-AP at Quadratic polynomial order

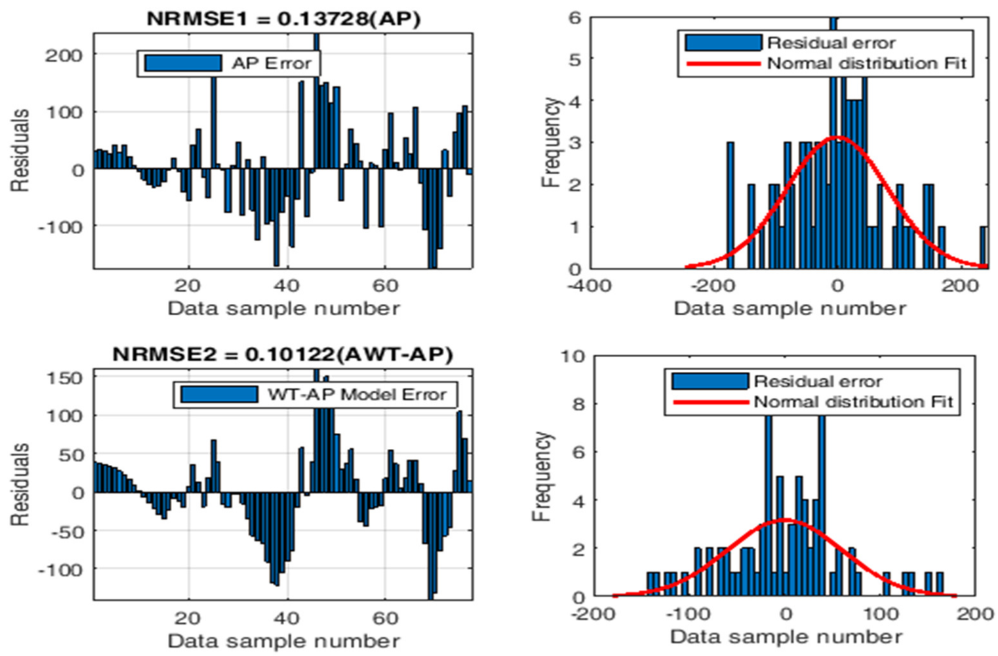

Fig. 7: Residual error plots of COVID-19 Infection rate trend estimation with AP and WT-AP at Cubic Polynomial order

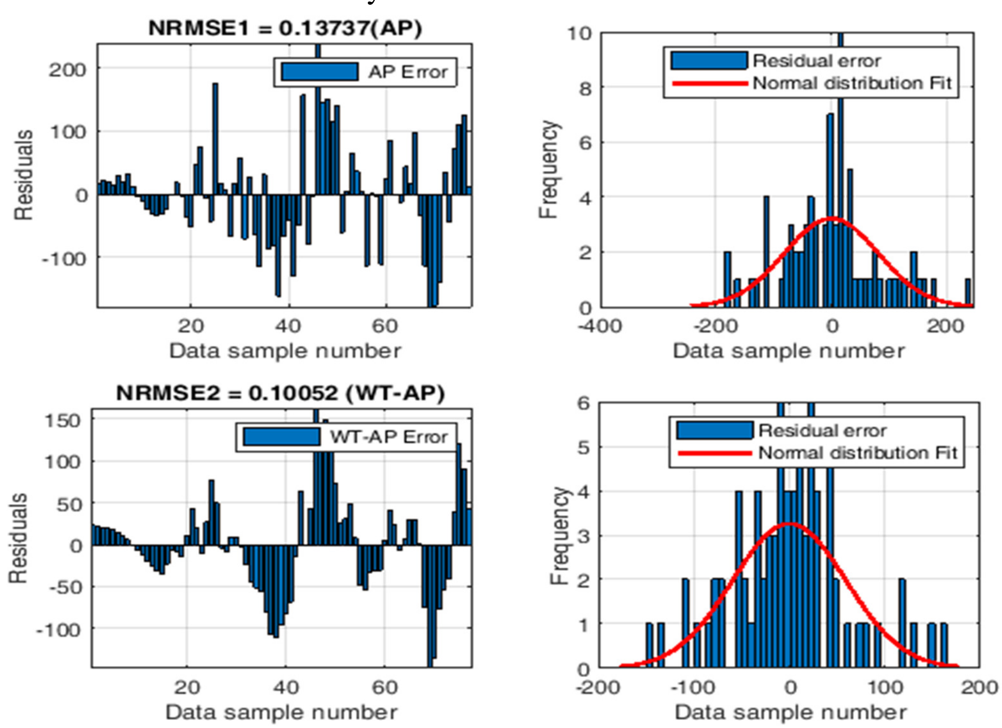

Fig. 8: Residual error plots of COVID-19 Infection rate trend estimation with AP and WT-AP at Quartic polynomial order

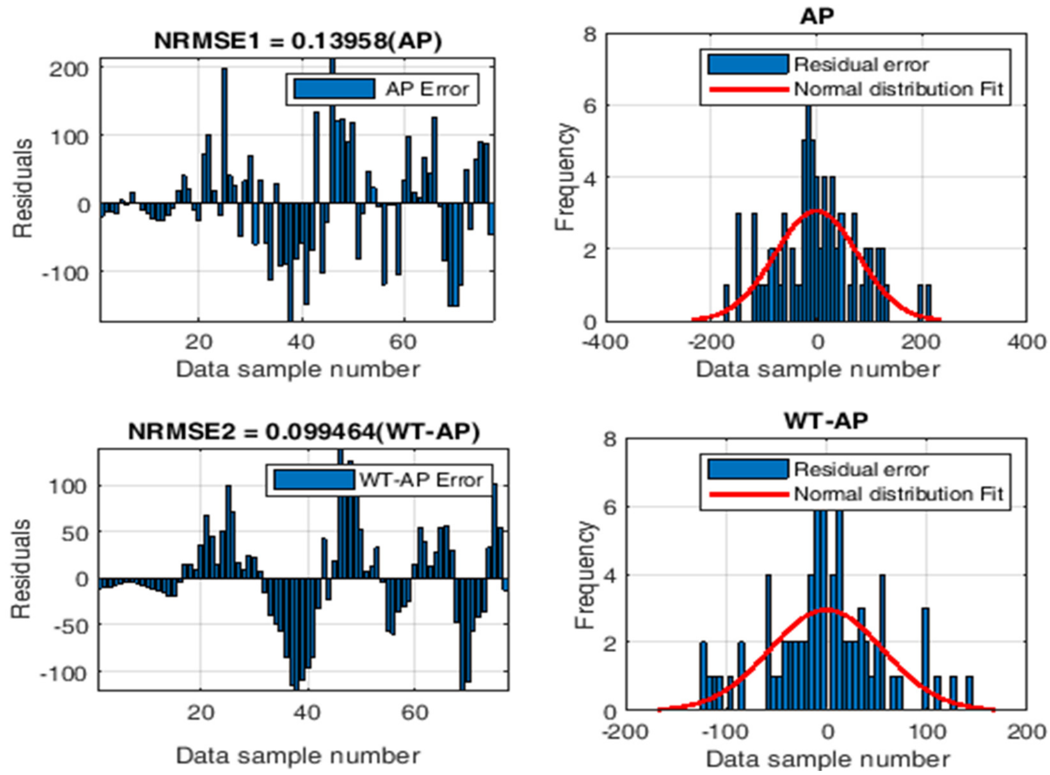

Fig. 9: Residual error plots of COVID-19 Infection rate trend estimation with AP and WT-AP at Quintic Polynomial order

Table 1: Parameter Estimates with AP and WT-AP at the different degree

Polynomial Degree order	Parameter Estimates(AP)	Parameter Estimates(WT-AP)
Quadratic	c(161.29 , 0.081928)	c(162.37 , 0.081095)
Cubic	c(-26.233 , 0.5443, -0.006354)	c(-27.229, 0.5486, -0.0064244)
Quartic	(-13.445, 0.45052, -0.0033498, -2.4192e-05)	c(-11.717, 0.43483, -0.0027801,-2.9346e-05)
Quintic	c(23.178, -0.1695, 0.030653, -0.00063983, 3.6264e-06)	c(22.575, -0.14574, 0.02906, -0.00060582, 3.3957e-06)

Table 2: P-values with AP and WT-AP at the different degree

Polynomial type	P-Values (AP)	P-Values(WT-AP)
Quadratic	1.807e-07, 2.0849e-11	4.2954e-08, 4.8069e-12
Cubic	0.17281,1.7531e-27, 9.2321e-24	0.05681,6.2618e-36, 7.2517e-32
Quartic	0.57769, 0.003349, 0.33627 0.38449	0.50749, 1.1172e-06, 0.27592 0.15126
Quintic	5.62e-240,0.37465, 0.007670, 0.0035127, 3.6264e-06	8.5422e-250, 0.286, 0.000509, 0.00015147, 0.0004676

Table 3: COVID-19 Estimation Accuracy with AP and WT-AP at the different degree

	Estimation Error Statistics (AP)			Estimation Error Statistics (WT-AP)		
Polynomial type	NRMSE	R^2	R^2(adjusted)	NRMSE	R^2	R^2(adjusted)
Quadratic	0.331	0.452	0.445	0.318	0.473	0.466
Cubic	0.317	0.861	0.858	0.301	0.919	0.917
Quartic	0.137	0.863	0.857	0.100	0.922	0.918
Quintic	0.139	0.872	0.867	0.009	0.93	0.927

CONCLUSION

This work proposed and explored a joint application of adaptive data trend modelling capacity of polynomial regression with broad data compression modelling strength of discrete wave transform to investigate and estimate the trend of coronavirus epidemic rate in Nigeria**.** The resultant statistical outcome shows that the most appropriate regression model to estimate COVID-19 epidemic data trend rate is the fifth degree polynomial regression model combined with wavelet transform scheme.

REFERENCES

Armstrong, B. (2006). Models for the relationship between ambient temperature and daily mortality. *Epidemiology,* 17:624–31.

Basu, R. (2009). High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008. *Environ Health*; 8:40.

Bell, M. L., Samet, J. M. and Dominici, F. (2004). Time-series studies of particulate matter. *Annu Rev Public Health* 25: 247–80.

Bhaskaran, K., Hajat, S., Haines, A., Herrett, E., Wilkinson, P. and Smeeth, L. (2010). Short term effects of temperature on risk of myocardial infarction in England and Wales: time series regression analysis of the Myocardial Ischaemia National Audit Project (MINAP) registry. *BMJ*;341:c3823.

Isabona, J. (2020). Wavelet Generalized Regression Neural Network Approach for Robust Field Strength Prediction, *Wireless Personal Communications,* 114(3):3635-3653.

Isabona, J., and Babalola, M. (2013). Statistical Tuning of Walfisch-Bertoni Pathloss Model based on Building and Street Geometry Parameters in Built-up Terrains. *American Journal of Physics and Applications*, 1:10-16.

Isabona, J. and Enagbonma, O. (2014). A Least Absolute Deviation Tuning Method to reduce Signal Coverage Loss Prediction Error in Electromagnetic Wave Propagation Channel optimization, *International Journal of Basic and Applied Sciences,* 1(1):44-55.

Isabona, J. and Isaiah, P. G. (2015). Computation and Verification of Propagation Loss Models based on Electric Field Data in Mobile Cellular Networks, *Australian Journal of Basic and Applied Sciences*, 9(29): 280-285.

Isabona, J., Konyeha, C. C., Chinule, C. B., Isaiah, G. P. (2013). Radio Field Strength Propagation Data and Pathloss calculation Methods in UMTS Network, *Advances in Physics Theories and Applications*, 21:54-68.

Jimenez, E., Linares, C., Martinez, D. and Diaz. J. (2010). Role of Saharan dust in the relationship between particulate matter and short-term daily mortality among the elderly in Madrid (Spain). *Sci Total Environ*, 408:5729–36.

Ogundokun, R. O., Lukman, A. F., Kibria, G. B.M., Awotunde, J.B. and Aladeitan, B. B. (2020). Predictive modelling of COVID-19 confirmed cases in Nigeria, *Infectious Disease Modelling,* 5:543-548.

Rath, S., Tripathy, A. and Tripathy, A. R. (2020). Prediction of new active cases of coronavirus disease (COVID-19) pandemic using Multiple Linear Regression Model, Diabetes and Metabolic Syndrome: *Clinical Research and Reviews*, 14:1467-1474.

Schwartz, J., Spix, C., Touloumi, G., Bacharova, L., Barumamdzadeh, T., leTertre, A., Peikarksi, T., Ponce de Leon, A., Ponka, A., Rossi, G., Saez. M. and Schouten, J. P. (1996). Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *J. Epidemiol. Community Health*, 50(Suppl.1):S3–11.

Qin, L., Sun, Q., Wang, Y., Wu, K. F. Chem, M., Sha, B. C. and Wu, S. Y. (2020). Prediction of number of cases 2019 Novel Coronavirus (COVID-19) using Social Media search Index, *Int J Environ Res Public Health*, 17(7): 2365.

Worldometer:https://www.worldometers.info/coronavirus/country/nigeria/(Accesed 4th August, 2018)