Examining the Efficacy of Break for Time Series Components (BFTSC) and Group for Time Series Components (GFTSC) with Volatile Simulated and Empirical Data.

Abstract:

The main reason for this study is to know the performance of BFTSC (break for time series components) and GFTSC (group for time series components) in identification of time series components using volatile simulated and empirical data. BFTSC was created to capture the trend, seasonal, cyclical and irregular components and presented them in a time series plot. While GFTSC was designed to capture all the four time series components together with the equations that produces each components of time series. BFAST (break for additive, seasonal and trend) only identifies trend and seasonal components while considering all other left over components as random, identification of trend and seasonal components alone is not enough to have a clear image of all the time series components in a time series data. Performance through evaluation using low and high volatile simulated and empirical data was conducted to evaluate the performance of both techniques. For yearly sample size of 8, 16 and 24 years were for small medium and large sample size. For the monthly data, 48, 96 and 144 months were used as small, medium and large sample size. Each of the sample size was replicated 100 times each. Finally, GFTSC and BFTSC performance was very good for large sample size with linear trend for both monthly and yearly data (approximately 100%). While the performance drops with highly volatile data such as trend with curve trend line (such as quadratic and cubic). These findings indicate that BFTSC and GFTSC can provide a better alternative to manual technique and BFAST for data associated with linear trend, hence BFTSC and GFTSC are recommended for public.

Keywords: Automation, Break for Time Series Components, Group for Time Series Components, Trend, Seasonal, Cyclical, Irregular.

Introduction

The purpose of this study is to evaluate the performance of BFTSC and GFTSC in identification of time series components. The technique BFAST was created to identify trend and seasonal time series components only while BFTSC was created for identification of the four time series components (such as trend, seasonal, cyclical and irregular time series components) and GFTSC was created to identify the four time series components together with the equations that produces each time series components (Ajare & Suzilah, 2019). Both GFTSC and BFTSC automated time series components identification were created from BFAST (Ajare & Suzilah, 2019). BFTSC and GFTSC are both improved BFAST. BFAST is a technique used for identification of trend and seasonal components of time series data, trend breaking was first suggested by Verbesselt et al. (2010) and was utilized by Jong, Verbesselt, Schaepman and Bruin (2012) recommended an approach of basic swing identification to spot time series component. This approach was also used by Zewdie, Csaplovics and Inostroza (2017) as the latest time series component recognition approach which is a technique that was first described and utilized by Verbesselt et al. (2010). The technique BFAST was for recognizing breaking points with the help of seasonal and trend decomposition using loess (STL), it facilitates the detection of trend change in a given information. The elementary aims of the BFAST technique is the splitting of time series into seasonal, trend and also remnants element by the approach for breaks detecting software in R studio core 2012 (Cunha, 2013).

Literature Review

Banas and Kożuch (2019) recommended the use of manual time series decomposition for identification of time series components in complex data such as variety of timber price and supply data. Multiplicative model was utilized as the product of four components while additive model is the addition of the four components (such as trend, seasonal, cyclical and irregular). The components of the time series were determined by means of the Census X11 technique but the cyclical component was detached from the trend by utilizing the Hodrick–Prescott filter (Banas & Kożuch, 2019).

Idrees, Alam and Agarwal (2019) described univariate time series components identification as a very important and vital tool for projection of data. Importantly, owing to its widespread use in various practical domains. Likewise stock market is considered to be one of the most highly volatile complex financial group which consist of various components or stocks, the price of which fluctuates greatly with respect to time. Stock market forecasting involves uncovering the market trends with respect to time. All the stock market investors aim to maximize the returns over their investments and minimize the risks associated. Stock markets being highly sensitive and volatile to quick changes, this suggested the use of ARIMA approach to be good enough for handling time series data and as such can be very constructive in various real world problems like that of health sector, education, finance and other practical domains for prediction. The main future study of stock-trend prediction is to develop new automated innovative model that can help to foresee the stocks result in high profits.

Sen (2019) utilized partial trend identification by change-point successive average methodology (SAM). There are diverse time series trend identification methodologies in the literature, but most of them are for monotonic trend explorations within a given hydro-meteorological time series. The most important question is how to identify the time series component successively, using trends of different durations and slopes and through automation. Successive arithmetic average methodology (SAM) is recommended, firstly, for a set of trends visual inspection and then their quantitative duration. The most important and vital advantage of SAM is that preliminary assumptions is not necessary and trends identifications are not straightforward (Sen, 2019). The limitation of SAM is that it is time consuming.

Xu, Yu, Peng, Zhao, Cheng, Liu and Gong (2020) examined the annual 30-m land use/land cover maps of China for 1980 until 2015. The length of each sample size was recommended to be within 5 to 15 years for breaks to be detected by the different break test approaches. Annual land use land cover (LULC) change information at medium sample size resolution is essential subjects for object model identification. Annual LULC observations is not always available at continental national scale due to insufficient remote sensing information coverage and lack of computational capabilities. Moderate Resolution Imaging Spectro radiometer (MODIS) and Global Inventory Modelling and Mapping Studies (GIMMS) normalized difference vegetation index (NDVI)) with high spatial resolution datasets (China's Land-Use/cover Datasets (CLUDs) were derived from 30-meter Landsat TM/ETM+OLI to generate reliable annual nominal 30 m LULC maps for the whole of China. The performance of a statistical test is based on change detection algorithm (Breaks for Additive Seasonal and Trend). Annual mapping results provided a detail process of urbanization, deforestation, afforestation, and water and cropland dynamics for over the past 35 years. The consistent characterization of land change dynamics for China can be further be studied using more years (16 years and above) for scientific research and to support land management for policy-makers.

Verbesselt, Hyndman, Zeileis and Culvenor (2010) used BFAST to examine the vegetation change and also accounting for abrupt and gradual trends in satellite image time series. This was done using diverse land cover types and seasonal amplitude to determines the signal-to-noise ratio. Utilizing the technique on 16-day NDVI MODIS images from 2000 until 2009 (10 years) for a vegetation research area. It was shown that a minimum seasonal amplitude of 0.1 NDVI is required to detect phenological change within cleaned MODIS NDVI time series using the quality flags. BFAST detect the phenological change independent of vegetation metrics by exploiting the full time series. The technique is globally applicable since it analyzes each pixel individually without the setting of thresholds to detect change within a time series. Long term vegetation changes can be detected within NDVI time series of a large range of land cover types (e.g. grassland, woodlands and deciduous forests) having a seasonal amplitude larger than the noise level. The technique can be applied to any time series data and it is not necessarily limited to NDVI data.

Verbesselt, Hyndman, Zeileis and Culvenor (2010) recommended that BFAST is capable of detecting and grouping spatial, temporal vegetation changes. BFAST is not specific to a particular data type and can be applied to time series without the need to normalize it for land cover types, or to select a reference period, or change trajectory. The technique can be integrated within monitoring frameworks and utilized as an alarm system to detect when and where changes occur. BFAST steps are iterated until the number and position of the breakpoints are detected.

Sang, Wang and Liu (2014) investigated the performances of two techniques of time series components identification. Mann–Kendall (MK) technique and the empirical model decomposition (EMD) technique for time series components identification in hydrological time series. Identification of time series components was a vital issue in hydrological time series study, but it was also very challenging task due to the diverse performances of different techniques. Examination of both synthetic and observed time series components identification techniques indicated a better performance of EMD compared with the MK technique.

The outcomes confirmed that pre-whitening cannot really improve time series component identification when using the MK technique, but produced non accurate results sometimes. If the trend time series component is analyzed with small magnitude of other time series components, it cannot be accurately identified by the MK technique, because the trend would not be submerged too severely by other time series components but trend component would be accurately identified. When analyzed series has short length, its trend cannot be accurately identified by the MK technique (Sang, Wang & Liu, 2014). However, the EMD technique can eliminate the influences of trends' magnitude and series' length, so it has more effective power for trend time series component identification. As a result, it is suggested that series' trend can be directly identified by the MK technique but need not pre-whitening; besides, the influences of trends' magnitude should be carefully considered for trend time series component identification. Comparatively, the technique can adaptively determine the specific shape of the nonlinear and non-stationary trend time series component by considering statistical significance, so it can be an effective alternative for time series components identification for hydrological time series.

Bonakdari, Moeeni, Ebtehaj, Zeynoddin, Mahoammadian and Gharabaghi (2019) investigated a new insights into soil temperature, time series modeling in linear and nonlinear. The spectral analysis technique is utilized, time series data from the time domain is transmitted to the new linear data and presented in the methodology. A methodology was proposed based on stochastic time series technique for modeling Daily Soil Temperature (DST). According to a comparison of the three techniques applied to analyze the various time series data, it appears that spectral analysis technique combined with stochastic technique outperformed the seasonal standardization and seasonal differencing technique.

Ambrosino, Thinová, Briestenský and Sabbarese (2019) examined the anomalies in identification of time series components in earth's rotation rate. The selected hybrid techniques are Empirical Mode Decomposition (EMD), Support Vector Regression (SVR), Singular Spectrum Analysis (SSA) and Forecasting Methodology (FM). Both hybrid techniques combined to form the first part of signal decomposition and a second part of signal modeling. The hybrid technique are formed by the combination of different techniques. The selected earthquakes period occurrence are estimated to show a direct consistency of hybrid technique. The EMD + SVR techniques has been proven to be the best for non linear time series data.

Parmezan, Souza and Batista (2019) examined the performance of statistical and machine learning models for time series modeling, identifying the best conditions for the use of each model. The choice of the most capable algorithm to model a particular phenomenon is one of the most vital events of univariate time series modeling. Modeling are similar to other data mining tasks, uses empirical evidence to select the most suitable model for a current problem since no modeling technique can be considered as the best. Only few technical research publications rigorously focus on the benefits and limitations of the most common algorithms for univariate time series processing. However, there are limited performance record of these models when applied to complex and highly nonlinear data. The outcomes indicated that SARIMA is the only statistical technique that is able to outperform other techniques in terms of time series trend component identification. Though, without a statistical difference machine learning algorithms such as artificial neural network (ANN), support vector machine (SVM) and Kth neural network (KNN) accuracy comes at the expense of a larger number of parameters. Detailed results achieved by different indexes as MSE, Theil’s U coefficient, POCID, are recently proposed multi-criteria performance measure are available online in our repository. The findings reveals that they helps in providing a broad insight into time series models selection, time series parameters setting, evaluation measures and experimental setup.

Awty, Bunting, Hardy and Bell (2019) investigate the performance of four time series components identification techniques using simulated univariate time series data. Based on the work done by Verbesselt et al. (2010) the commonly simulated data ranges from 5 years until 24 years. 10 years’ time series data that contained a 16-day temporal resolution gives approximately 23 observations per year and curve is centered on the middle of the year. The noise component was included randomly to create more realistic time series data.

Evaluating the correctness and limitations of these techniques can be difficult because validated data are limited and commonly depends on human interpretation. Data generated through time series simulation offer an objective technique for comparison between change detection algorithms. In total, 151,200 monthly data simulations were generated to represent a range of abrupt, gradual, and seasonal changes. Exponentially Weighted Moving Average Change Detection (EWMACD) and Break for Additive, Seasonal and Trend (BFAST) performed very well. EWMACD correctly identified the true date of change in 76.6% of cases. Continuous Change Detection and Classification (CCDC) and Continuous Change Detection and Classification plus Cross Validation (CCDC+CV) performed worst. BFAST Monitor improved when data were removed. Though BFAST could only correctly identify less than 10% of seasonal changes but 100% of linear trend. All in all the techniques showed some decrease in performance with increased noise (with highly volatile data) and missing data. The following recommendations are made from the limitations of each techniques as a starting point for future studies. EWMACD should be utilized for detection of lower magnitude trend changes and changes in seasonality. CCDC should be used for robust detection of complete land cover class changes. EWMACD and BFAST are suitable for noisy datasets and they are both recommended as the best time series components identification technique. BFAST can be extended to detect cyclical and irregular components in addition to linear trend and seasonal components. CCDC should be used where there are high quantities of missing data. The simulated datasets have been made freely available online as a foundation for future work.

Verbesselt, Hyndman, Zeileis and Culvenor (2010) recommended BFAST technique for public use. BFAST has mainly been applied to monitoring forest disturbance but can also be applied to more general land cover monitoring scenarios. First, an Ordinary Least Squares Moving Sum (OLS-MOSUM) test is used to determine if any breakpoints are present in the time series. If the OLS-MOSUM test indicates significant ($p < 0.05$) change, the number and location of breakpoints is estimated separately for the seasonal and trend components using OLS fitting. The BFAST package automatically fits a third-order harmonic model. The result is a set of piecewise season-trend models which minimise error across the whole time series. The minimum distance between breaks was set to 2 to 6 years (46 observations), which is in line with the guidelines given by Verbesselt et al. (2010) and matches the two-training period that was used for the other techniques.

BFAST Monitor was developed as a near-real time alternative to BFAST (Verbesselt, Zeileis and Herold (2012). BFAST Monitor is similar to BFAST, it was mainly been applied to forest monitoring. It is built on the fact that change can be detected by looking for deviation of new observations from an established data history period. BFAST Monitor is not built for separating seasonal and trend components. The season-trend model is fitted to the stable history data period using OLS. Whenever new observations are available, the residual values are calculated using the fitted model and Moving Sums (MOSUMs) of the residuals are used to look for instability which would indicate structural change. BFAST Monitor was run using the R package the same manner as BFAST is being run. Given that all simulations were designed with a break after five to six years of stability, a stable history period of two years (46 observations) was used. BFAST Monitor can be used on datasets with missing values which give it better advantage over BFAST. BFAST Monitor uses the difference in medians between the history period and monitoring period to estimate break magnitude. The basic limitation of BFAST monitor is the R implementation which does not allow for continuous monitoring and could not also identify cyclical and irregular time series components (Verbesselt, Zeileis and Herold, 2012).

Continuous Change Detection and Classification (CCDC) was built to primarily focus on univariate time series components identification in land cover time series (Verbesselt, Zeileis and Herold, 2012). Classification time series component was not used here because the simulated data were not designed to relate directly to specific land cover types. CCDC is similar to BFAST Monitor, CCDC focus on detecting changes in near-real time. The time series model used by CCDC is very similar to the season-trend model used by BFAST Monitor, except that CCDC uses an adaptive process to minimise model overfitting while also robustly capturing the seasonal cycle. Lasso Regression Model (LRM) instead of OLS to avoid overfitting of higher-order models to season-trend model to the historical period (Verbesselt, Zeileis & Herold, 2012). LRM reduces overfitting by limiting the total absolute value of the coefficients. As a result, some coefficients will be forced to zero and will have no influence on the model. There was no freely available implementation of CCDC suitable for use with simulated data so a suitable implementation was written in the Python programming language. Most of the time series components identification technique rely to some extent on parameter tuning to achieve the best results. The major limitation of CCDC is due to the use of Lasso fitting, CCDC is somehow dependent on the user to choose the numbers of harmonics or the length of the historical period and it is not consistent due to this facts that it is dependent on users.

CCDC with Cross validation was used to find the optimal value for fitting multiple models with different values and comparing them. These two approaches are referred to as CCDC and CCDC with Cross Validation (CV). For the fixed approach, a value 0.01 was chosen based on small scale testing. Some studies have reported values of 20. The time series models was fitted to surface reflectance rather than NDVI. CCDC with CV gave the third best performance among BFAST, BFAST monitor, CCDC, CCDC+CV and EWMACD (Deng & Zhu, 2020).

According to Deng and Zhu (2020), the performance of CCDC, CCDC with CV, and EWMACD were very similar at estimating trend break detection. This technique of break estimation is very robust to missing data but less effective with increased noise. The influence of noise is always affecting the performance of this techniques, this techniques relies on residual values, the more noisy the data, the less likely to reflect the true break size. One of the limitations of CCDC with CV is that it required much longer time to run than the other three techniques. Final findings reveals that using CV made CCDC can much more detect true breaks, but also has the probability of detecting at least one, false break in a time series data.

The technique BFAST had much lower RMSE and was more robust against noise, Hence BFAST is recommended as one of the best trend break detection. One of the limitation of CCDC with CV is that its algorithm was made complicated, unlike CCDC, CCDC with CV did not have a straightforward relationship between RMSE number of breaks and noise. CCDC with CV was also found to be less accurate (Zhu, Zhang, Yang, Aljaddani Cohen, Qiu & Zhou, 2020). Another limitation is also in terms of noise, with increased noise, the technique was less likely to detect correct results and the likelihood of detecting at least one false break remained constant. However, the change in RMSE tells us that the actual number of false breaks found is likely to be higher at extremes of noise. At high levels of noise, models are more likely to be influenced by noisy data points and may be fitting to noise. This is probably why most of the techniques performed less in terms of RMSE number of breaks at high noise levels. The unique pattern shown by CCDC with CV suggests that it must also detect more breaks if there is very little noise (Zhu, Zhang, Yang, Aljaddani Cohen, Qiu & Zhou, 2020).

EWMACD was built to focus on subtle changes, such as partial changes within pixels (Brooks, Wynne, Thomas, Blinn, Coulston, 2019). Just like CCDC and BFAST Monitor, EWMACD also detects condition (increasing/decreasing trend) changes because it only fits a seasonal model without a trend term. EWMACD uses a specific type of statistical control chart, the EWMA chart, to rapidly help in identification of time series component.

Zhu, Zhang, Yang, Aljaddani, Cohen, Qiu and Zhou (2020) developed a new univariate time series components identification method known as COntinuous Monitoring of Land Disturbance (COLD) using Landsat time series data. COLD can detect many time series component such as trend and seasonal. COLD can also detect land disturbance continuously as new pattern is collected and likewise provide historical land disturbance history. Evaluation of the trend detection ability and land disturbance, different kinds of data are utilized. The COLD algorithm was developed and calibrated based on all the lessons learned. The accuracy assessment shows that COLD results were accurate for detecting trend and seasonal as land disturbance with an omission error of 27% and a commission error of 28%. The limitation of COLD was inability to detect time series components accurately with large noise.

Statistical Control Charts (SCC) was developed as a form of univariate time series components identification method and used as control limits to establish when a time series data deviates from a stable state. The Moving Sum (MOSUM) and Cumulative Sum (CUSUM) charts used by BFAST and BFAST Monitor are other examples of statistical control chart. EWMACD calculates the residuals for a given training period based on a seasonal model fitted with OLS (Brooks, Wynne, Thomas, Blinn and Coulson, 2013). To match BFAST Monitor, a second-order seasonal model and a two-year history period were used. This produces a set of normally distributed, independent observations suitable for use with a EWMA chart. EWMACD has a free available version in goggle schola. The limitation was that the version does not allow for continuous monitoring and therefore we also drew from a later implementation of EWMACD called dynamic EWMACD. As of January 2015, BFAST was still one the most widely utilized time series components identification technique and is freely available online: https://cran.r-project.org/web/packages/bfast/bfast.pdf. The identification of time series components are summarised starting from 1960 to date. The strengths and weakness of each period and BFTSC and GFTSC was created (Ajare E.O. & Suzilah, 2019).

Material and Methods

BFAST is the technique used in identifying the time series data by extracting the trend and seasonal pattern during time series decomposition. Given the general time series additive model of the form:

$$Y_P = T_p + S_p + C_p + I_p \qquad (2.1)$$

Where Y_P is the observed value at time period p and T_p is the trend value at time period p, while S_p is the seasonal component value, C_p is the cyclical component and I_p is the irregular component all with time period p (Maggi, 2018 ; Zhao, Li, Mu, Wen, Rayburg, & Tian, 2015).

From equation (2.1) BFAST takes all other components relatively trend and seasonal component to be randomized (R_p) and the equation was expressed as

$$Y_P = T_p + S_p + R_p \qquad (2.2)$$

(Zdravevski, Lameski, Mingov, Kulakov & Gjorgjevikj, 2015).

To generate trend components using BFAST, we need a piecewise linear model approach. Suppose T_p is a piecewise linear model with an actual slope and intercept on q+1 segments broken with q breakpoints and P period; $p_1^{\neq},\ldots\ldots\ldots, p_q^{\neq}$ then T_p can takes the form

$$T_p = \alpha_k + \beta_k \mathrm{P}$$

Where $p_{k-1}^{\neq} < \mathrm{p} \leq p_k^{\neq}$

And If k = 1,…..,q then $p_0^{\neq} = 0$ and $p_{q+1}^{\neq} = \mathrm{n}$.

The slope of the change before the breakpoints while β_{k-1} and the slope of the breaks after the change breakpoints are β_k. The intercept and the slop of the linear model α_k and β_k with time period p and it will be used to derive the magnitude and direction of change.

To generate seasonal components using BFAST, we need a simple harmonic model.

Thus, S_p can be represented by a simple harmonic model with j terms; j = 12….J and time t.

$$S_p = \sum_{j=1}^{J} \omega_{k,j} \ \mathrm{Sin}\left(\frac{2\pi jt}{F} + \sigma_{K,j}\right) \qquad (2.3)$$

Where k = 1… q, $p_{k-1}^{\neq} < \mathrm{p} \leq p_k^{\neq}$ and also $\omega_{k,j}$, $\sigma_{K,j}$ are the segment amplitude and F is the frequency (Zeileis, Kleiber, Krämer & Hornik, 2003).

To generate random components, any data that does not belong to trend nor seasonal is classified randomR_p.

$$Y_P = \underbrace{\{ \alpha_k + \beta_k P\}} + \underbrace{\left\{ \sum_{j=1}^{J} \omega_{k,j} \ \ \text{Sin} \left(\frac{2\pi jt}{F} + \sigma_{K,j} \right) \right\}} + \underbrace{R_p} \quad (2.4)$$

$$Y_P = \quad T_p \quad + \quad S_p \quad + \quad R_p$$

According to Ajare and Suzilah (2019) the new technique called BFTSC and GFTSC considered splitting the random into cyclical components and irregular components which is an extension of BFAST. Cyclical components can be calculated through the regression cyclical movement. The regression function at the breakpoint maybe discontinuous but the model can be written in such a way that the function continues at all point including breakpoints. To calculate cyclical components, center moving average is involved (Bornhorst, Dobrescu, Fedelino, Gottschalk & Nakata, 2011).

The new equation becomes

$$Y_P = \underbrace{\{ \alpha_k + \beta_k P\}} + \underbrace{\{ \sum_{j=1}^{J} \omega_{k,j} \ \ \text{Sin} \left(\frac{2\pi jt}{F} + \sigma_{K,j} \right)\}} + \underbrace{\{ \frac{CMA}{\hat{CMA}} \}} + \{ \underbrace{I_p} \} \quad (2.5)$$

$$Y_P = \quad T_p \quad + \quad S_p \quad + \quad C_p \quad + \quad I_p$$

(Ajare and Suzilah, 2019)

Where Y_P is the observed value at time period p and T_p is the trend value at time period p, while S_p is the seasonal component value, C_p is the cyclical component and I_p is the irregular component at period p. I_p is the remainder variations which is not captured by trend, seasonal variations and cyclical components (I_p).

BFTSC and GFTSC technique considers every vital component of time series (Ajare & Suzilah, 2019). BFAST is known to be weak in identifying and breaking random components, also very weak in applicability to other types of empirical data (Flicek & Birney 2009). The problem of time series components detection is a problem that should be solved in the earliest stage of time series forecasting (Flicek & Birney 2009). BFTSC followed similar derivative steps like BFAST but deviated in the addition of cyclical and irregular components. BFTSC is the technique used in analyzing the generality of time series data by extracting the trend, seasonal, cyclical and irregular components during time series decomposition. GFTSC followed similar derivative steps like BFTSC but in addition to identification of trend, seasonal, cyclical and irregular components GFTSC also have the capability of generating the equations that produces each component. The residual component in BFAST now converted to contained cyclical and irregular component in GFTSC and BFTSC. Both BFTSC and GFTSC are automated time series components identification. In BFAST only random component can be observed but in BFTSC & GFTSC, the cyclical and irregular components were included (Ajare & Suzilah, 2019).

Data Simulation

This study compare the performance of BFTSC and GFTSC using both simulated and empirical data. In the simulation study, monthly and yearly data were replicated 100 times based on 3 sample sizes (small, medium and large) and by embedding the four time series components as the simulation conditions. Percentages were calculated in identifying the correct time series components that existed in the simulation. Simulation of 8 16 and 24 years were used as year simulated sample sizes. Simulation of 48, 96 and 144 months were used for monthly sample sizes. Three types of trends were used: linear, quadratic and cubic for both yearly and monthly simulations. The data were replicated 100 times for each condition percentages of identifying correct time series components were computed to evaluate BFTSC and GFTSC respectively. Each of the simulated set of data contains component combinations of different form; where monthly data has Trend (T_t), Trend and Seasonal (T_tS_t), Trend and Irregular (T_tI_t), Trend and Cyclical (T_tC_t), Trend, Seasonal and Irregular ($T_tS_tI_t$), Trend, Seasonal and Cyclical ($T_tS_tC_t$), Trend, Seasonal, Irregular and Cyclical ($T_tS_tI_tC_t$). As for yearly data has Trend (T_t), Trend and Irregular (T_tI_t), Trend and Cyclical (T_tC_t), Trend, Irregular and Cyclical ($T_tI_t\ C_t$). Evaluation of BFTSC and GFTSC is based on the ability to identify the correct time series components in 100 replications. Though the issue of how large is large and maximum data accepted by BFAST is yet to be addressed (Van Leeuwen, Huete and Laing, 1999).

Table 1 lists the equations and conditions for monthly and yearly data simulation. First, three different types of trend (linear, quadratic and cubic) were generated randomly based on different values of the coefficient a, b, c, d by using the time variable (t) to replicate 100 set of data for trend component condition. Next, the trend equations were adjusted accordingly by adding other components conditions which were seasonal, cyclical and irregular for monthly data; and cyclical and irregular for yearly data. The seasonal, cyclical and irregular values were obtained from the average, maximum and double maximum values respectively, from the trend randomly generated data. Then, the data was simulated by incorporating 12 sets of seasonal, 3 sets of cyclical, 2 sets of irregular and embedded in 3 sample sizes of small, medium and large of monthly (48, 96 and 144 months respectively) and yearly data (8, 16 and 24 years respectively). The seasonal adjustment was implemented by deducting average value to the trend values at 12 places. As for cyclical, maximum value was added to the trend values at 3 places and for irregular, double maximum value was added at 2 places. Table 1 displays the strategy use in generating the trend (with linear, quadratic and cubic) for yearly and monthly time series data.

Table 1

Monthly and Yearly Data Simulation (Equations and Conditions)

Time Series Components	Types of Trend	Equations
Trend (T_t)	Linear	$T_t = a + bt$ $100 \leq a \leq 200$, $201 \leq b \leq 300$
	Quadratic	$T_t = a + bt + ct^2$ $100 \leq a \leq 200$, $201 \leq b \leq 300$, $301 \leq c \leq 400$
	Cubic	$T_t = a + bt + ct^2 + dt^3$ $100 \leq a \leq 200$, $201 \leq b \leq 300$, $301 \leq c \leq 400$. $401 \leq d \leq 500$
Seasonal (S_t)	Adjusted at 12 places using average value of the trend generated data	
Cyclical (C_t)	Adjusted at 3 places using maximum value of the trend generated data	
Irregular (I_t)	Adjusted at 2 places using double maximum value of the trend generated data	
Overall Equation	$Y_t = T_t + S_t + C_t + I_t$ (monthly) $Y_t = T_t + C_t + I_t$ (yearly)	

The main model used in generating subsequent trend data, the data was adjusted to add other components appropriately. The first trend data is in one hundred replicates involving only trend component. The second set of data, 12 seasonal components were added to the trend to form trend and seasonal. The third set of data, 2 irregular components was added to the trend to form trend and irregular. The fourth set of data, 3 cyclical components was added to the trend to form trend and cyclical and the last set of data is the combination of the four time series components. The same approach above will be use in generating other subsequent monthly trend data for 96 months and 144 months, such that they all contain only trend and are replicated in 100 places.

Evaluation using simulated data

Evaluation of both techniques is the process of examining the efficiency of BFTSC and GFTSC using data generated from the equations in Table 1 through the simulation study. Table 2 and 3 display the simulation results based on linear trend as the basis for time series components combinations using both monthly and yearly data respectively. Both BFTSC and GFTSC performed very well in the large sample size (144 months) together with all different conditions of time series components. Both techniques successfully identified 100% of the correct components (Table 2). However, the small and medium sample sizes have percentages of 81% and 90%, respectively, when the combinations became more complex (Linear Trend, Seasonal and Cyclical ($T_tS_tC_t$); and Linear Trend, Seasonal, Irregular and Cyclical ($T_tS_tI_tC_t$)). As for the yearly simulated data, BFTSC managed to obtain 100% correct identification for only two conditions which were Linear Trend (T_t), and Linear Trend and Irregular (T_tI_t) for all sample sizes (small (8 years), medium (16 years) and large (24 years)) as shows in Table 3. However, when the conditions were Linear Trend and Cyclical (T_tC_t), and Linear Trend, Irregular and Cyclical ($T_tI_tC_t$) with small (8 years) and medium (16 years) sample sizes; the performance of BFTSC deteriorated 2% and 1% respectively. This is due to the small and medium sample sizes of 8 and 16 data points in capturing the complexity of the time series components conditions where involving not just linear trend but 3 sets of cyclical and 2 sets of irregular.

Table 2. Evaluation of BFTSC and GFTSC with linear trend using small, medium and large sample size (months)

	Percentages of correct identification (%)		
Time Series Components	***Small (48 months)***	***Medium (96 months)***	***Large (144 months)***
Linear Trend (T_t)	*100 %*	*100 %*	*100 %*
Linear Trend and Seasonal (T_tS_t)	*100 %*	*100 %*	*100 %*
Linear Trend and Irregular (T_tI_t)	*100 %*	*100 %*	*100 %*
Linear Trend and Cyclical (T_tC_t)	*100 %*	*100 %*	*100 %*
Linear Trend, Seasonal and Irregular ($T_tS_tI_t$)	*100 %*	*100 %*	*100 %*
Linear Trend, Seasonal and Cyclical ($T_tS_tC_t$)	*81 %*	*90 %*	*100 %*
Linear Trend, Seasonal, Irregular and Cyclical ($T_tS_tI_tC_t$)	*81 %*	*90 %*	*100 %*
Averages	*95 %*	*97 %*	*100 %*

Table 3. Evaluation of BFTSC and GFTSC with linear trend using small, medium and large sample size (years)

	Percentages of correct identification (%)		
Time Series Components	***Small (48 months)***	***Medium (96 months)***	***Large (144 months)***

Trend (T_t)	*0 %*	*0 %*	*0 %*
Trend and Seasonal (T_tS_t)	*50 %*	*55 %*	*58 %*
Trend and Irregular (T_tI_t)	*0 %*	*0 %*	*0 %*
Trend and Cyclical (T_tC_t)	*0 %*	*0 %*	*0 %*
Trend, Seasonal and Irregular ($T_tS_tI_t$)	*33 %*	*36 %*	*38 %*
Trend, Seasonal and Cyclical ($T_tS_tC_t$)	*33 %*	*36%*	*38 %*
Trend, Seasonal, Irregular and Cyclical ($T_tS_tI_tC_t$)	*25 %*	*26%*	*28 %*
Averages	*20 %*	*22 %*	*23 %*

Table 4 Evaluation of BFTSC and GFTSC with Quadratic and Cubic trend using small, medium and large sample size (months)

	Percentages of correct identification (%)		
Time Series Components	***Small (48 months)***	***Medium (96 months)***	***Large (144 months)***
Trend (T_t)	*0 %*	*0 %*	*0 %*
Trend and Seasonal (T_tS_t)	*50 %*	*55 %*	*58 %*
Trend and Irregular (T_tI_t)	*0 %*	*0 %*	*0 %*
Trend and Cyclical (T_tC_t)	*0 %*	*0 %*	*0 %*
Trend, Seasonal and Irregular ($T_tS_tI_t$)	*33 %*	*36 %*	*38 %*
Trend, Seasonal and Cyclical ($T_tS_tC_t$)	*33 %*	*36%*	*38 %*
Trend, Seasonal, Irregular and Cyclical ($T_tS_tI_tC_t$)	*25 %*	*26%*	*28 %*
Averages	*20 %*	*22 %*	*23 %*

Table 5. Evaluation of BFTSC and GFTSC with Quadratic and Cubic trend using small, medium and large sample size (years)

	Percentages of correct identification (%)		
Time Series Components	***Small (8 years)***	***Medium (16 years)***	***Large (24 years)***
Trend (T_t)	*0 %*	*0 %*	*0 %*
Trend and Irregular (T_tI_t)	*0 %*	*0 %*	*0 %*
Trend and Cyclical (T_tC_t)	*0 %*	*0 %*	*0 %*
Trend, Irregular and Cyclical ($T_tI_tC_t$)	*0 %*	*0 %*	*0 %*
Averages	*0 %*	*0 %*	*0 %*

Table 4 and 5 show the simulation results of both techniques based on monthly and yearly data generated from quadratic and cubic trends, respectively. The results were the same for both trends (quadratic and cubic); thus, they were combined. BFTSC and GFTSC performed poorly in quadratic and cubic trends for monthly data. They were only able to identify the time series components on average of 20% to 23% for all sample sizes (small (48 months), medium (96 months), and large (144 months)) as indicated in Table 4. Both techniques (BFTSC and GFTSC) also failed to identify any time series components in yearly data (0%) for all sample sizes, either small (8 years), medium (16 years), or large (24 years), as revealed in Table 5. These poor performances were due to the derivation of BFTSC and GFTSC was based on BFAST linear trend only and not other types of trends.

Both BFTSC and GFTSC performed very well (100%) in the simulation study of linear trend with different combination of time series components (seasonal, irregular and cyclical), and large sample size (144 months and 24 years) for monthly and yearly data. However, it was noticeable that the performances were affected by small and medium sample sizes with complex combination of time series components of linear trend for monthly and yearly data. The correct identification percentages decreased by 19% and 2% for small sample size and 10% and 1% for medium sample size for monthly and yearly data respectively. This is due to the limited number of data points (i.e., monthly: 48 and 96; yearly: 8 and 16 only) used in identifying the complex combination of linear trend, 12 sets of seasonal, 3 sets of cyclical and 2 sets of irregular. Nevertheless, BFTSC and GFTSC performed poorly when quadratic and cubic trend were used. This is due to the derivation of BFTSC and GFTSC was based on linear trend as in BFAST. Thus, to use BFTSC and GFTSC requires large sample size to identify the existence of linear trend and other time series components (seasonal, irregular, and cyclical) in a data set.

Evaluation using Empirical data

Figure 1 displays the time series plot of Ibadan monthly rainfall from January 2007 until December 2018 using automated BFTSC. A total of 144 months, it was noticeable they were regular repeated patterns across the months and years. The high amount of rainfall was between May until August and low amount of rainfall between November until March due to wet and dry season respectively. This is because of the geographical location of Ibadan near to the lagoon. The plot (Figure 1) also shows slight increment in the regular fluctuation starting from 2011 onwards. These regular patterns closely related to seasonal component.

Figure 1 shows the plots produced by automated BFTSC when identifying the time series components in Ibadan rainfall data. BFTSC combined the 4 plots simultaneously for easy, straight forward and fast identification process. Figure 2 is the second plot which was the automated GFTSC plot of actual monthly rainfall data; this give similar representative of the Ibadan monthly rainfall as BFTSC in figure 2. Figure 3 is the manual time plot of Ibadan monthly rainfall. These findings were the same as manual identification approach in plot 3. GFTSC separated the components and produced equation or time series component values on top of each plot.

The identification results obtained by GFTSC were the same as BFTSC because of using the same theoretical derivation as detailed earlier . Thus, GFTSC findings were also the same as the manual identification approach. The difference between BFTSC and GFTSC was in producing the plots where GFTSC provides more details by including the equation and time series component values which can be used for deeper understanding of the data and also for forecasting. Both BFTSC and GFTSC are more convenient to use by end users because of easy, straight forward and fast identification process.

Figure 3 presents the manual time series plot of Ibadan yearly total rainfall from 2007 until 2018. This was a confirmation of BFTSC and GFTSC efficiency in identification of time series components. Now, it was much clearer, there was trend and also showing a linear trend which was very difficult to detect in Figure 3 above due to seasonal fluctuation. The amount of rainfall 2007 until 2009 was lower because of La Nina effect and higher in 2010 to 2011 due to El Nino. Then the rainfall decreased in 2012 was also due to La Nina. However the increment of rainfall in 2013 was due to global warming (Nucitteli, 2014). As for 2016, the high amount of rainfall was due to El Nino (Null, 2021).

FIGURE 1. BFTSC OF 10 YEARS MONTHLY RAINFALL IN IBADAN

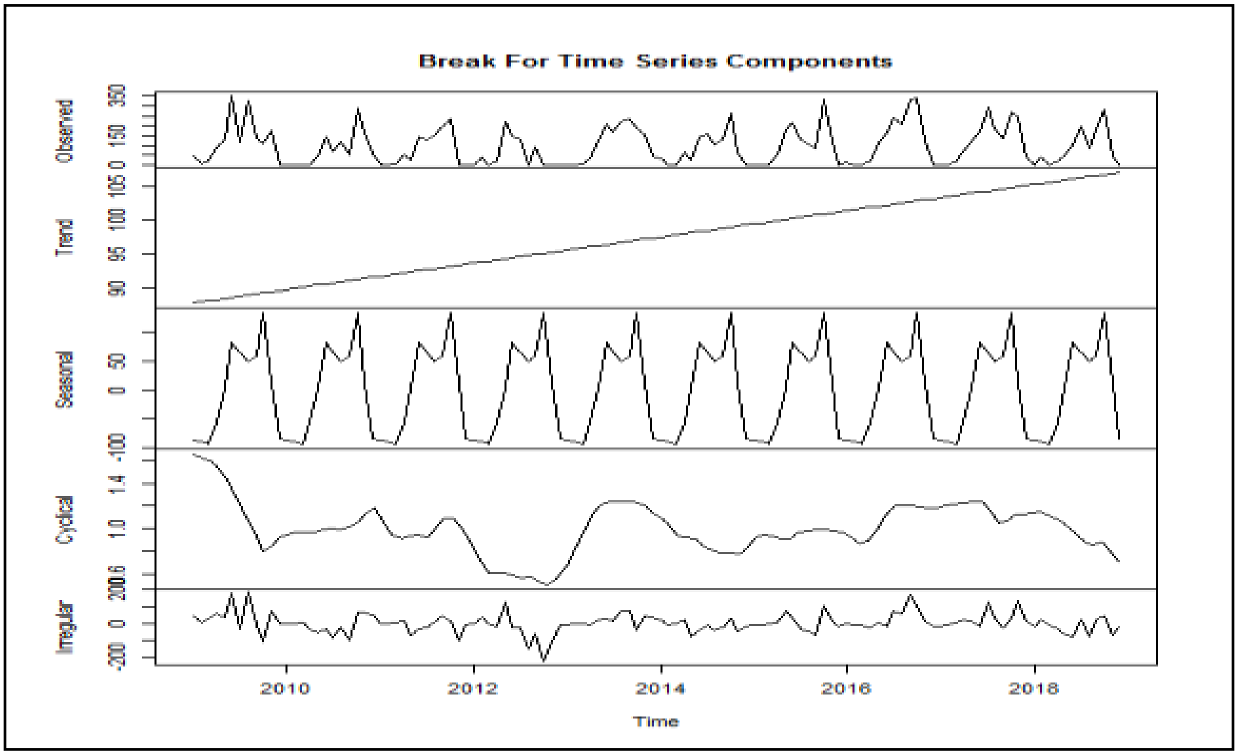


FIGURE 2. GFTSC OF 10 YEARS MONTHLY RAINFALL IN IBADAN

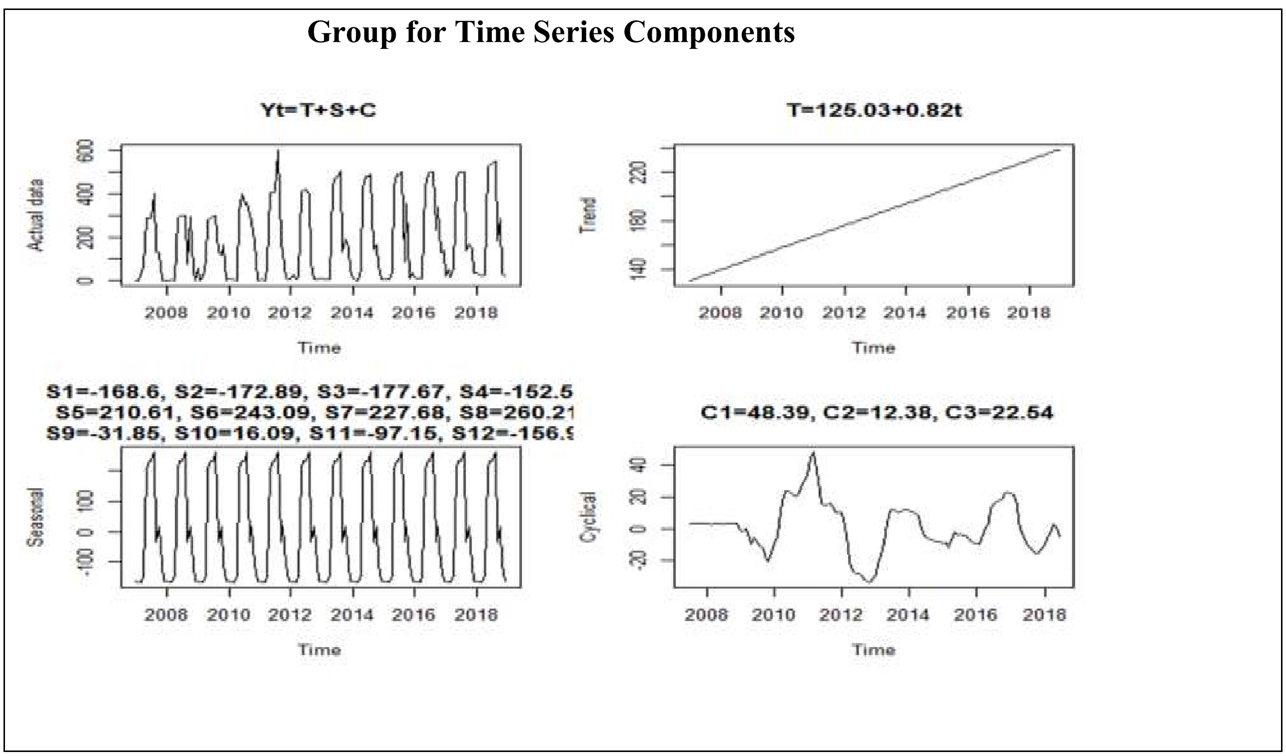
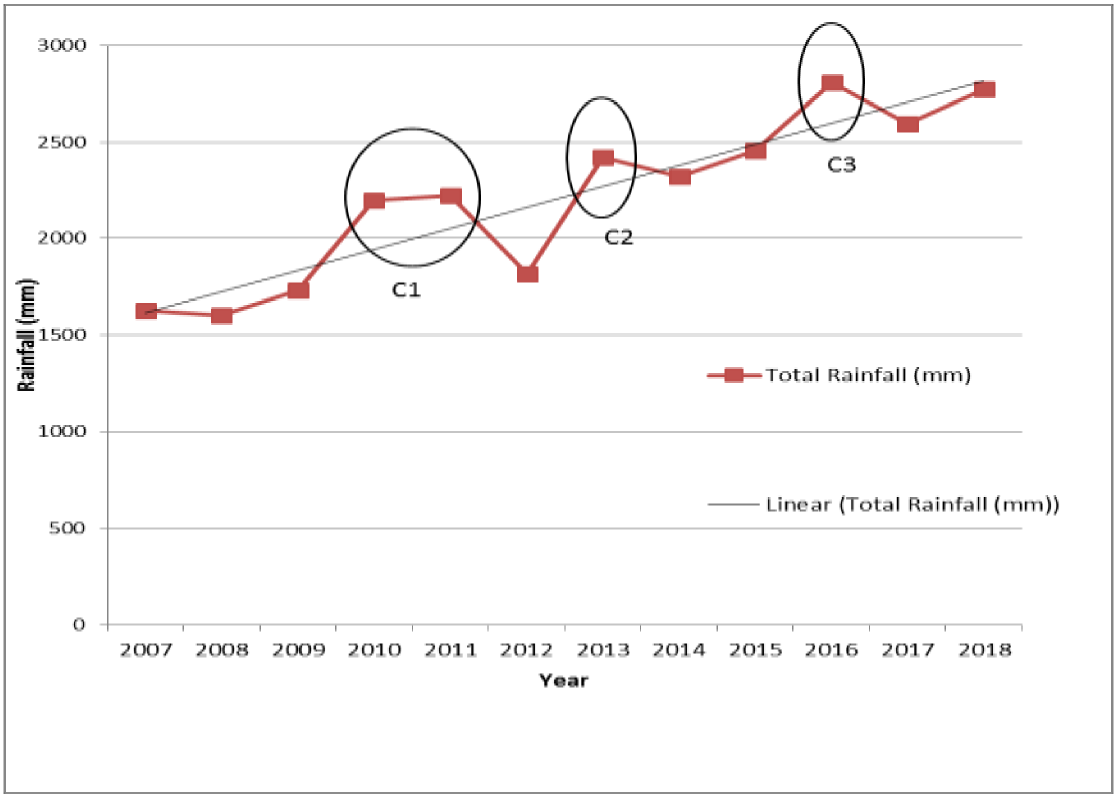


Figure 3. TIME SERIES PLOT OF IBADAN YEARLY TOTAL RAINFALL



BFTSC and GFTSC using UK GDP

Figure 4, 5 and 6 show the plots produced by BFTSC, GFTSC and time series plot for UK GDP yearly data respectively. BFTSC display simultaneously the plots of actual UK GDP, trend and cyclical as in Figure 4 which was easy, straight forward and fast in identifying the existence of time series components in the data. GFTSC also display the 3 plots but together with the trend equation and cyclical value (C1). This can enhanced further understanding regarding the UK GDP data and also can be used in forecasting. Both BFTSC and GFTSC successfully extract trend and cyclical components as identified by manual approach as in previous section

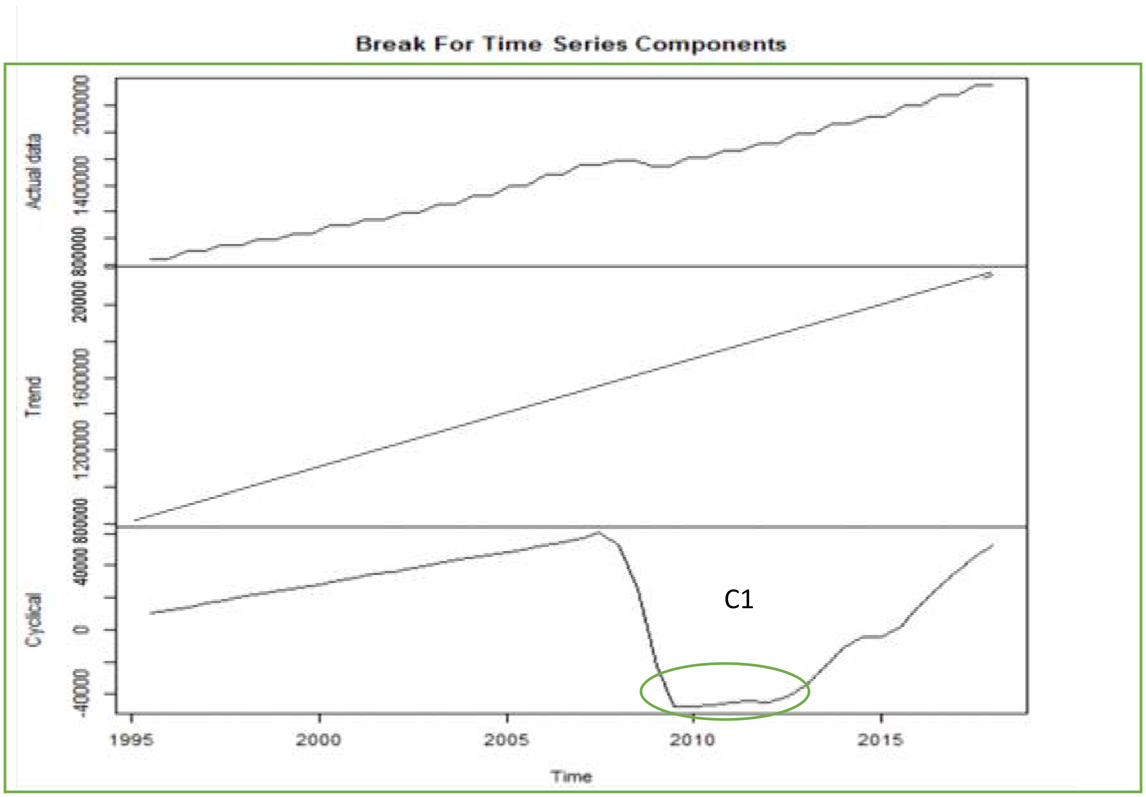


Figure 5 GFTSC plots of Yearly UK GDP

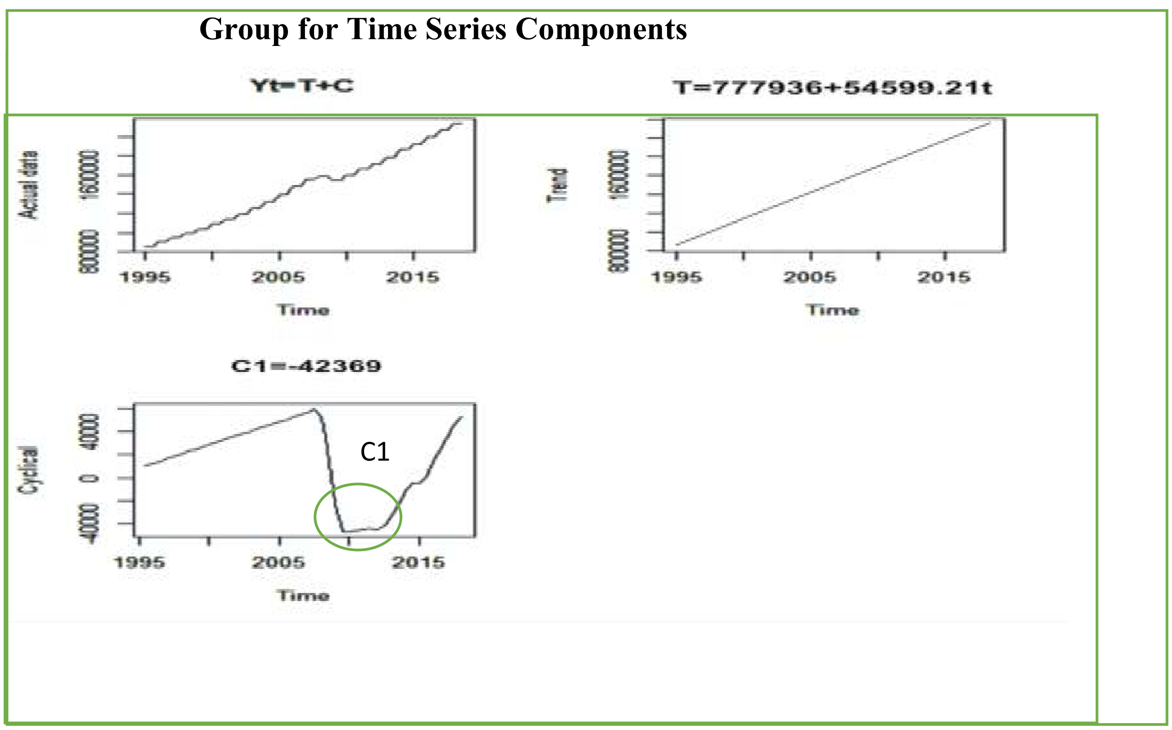


Figure 6. UK GDP Yearly TIME SERIES PLOT

Fig. 7. United Kingdom Gross Domestic Product (UK GDP) from 1995 to 2018

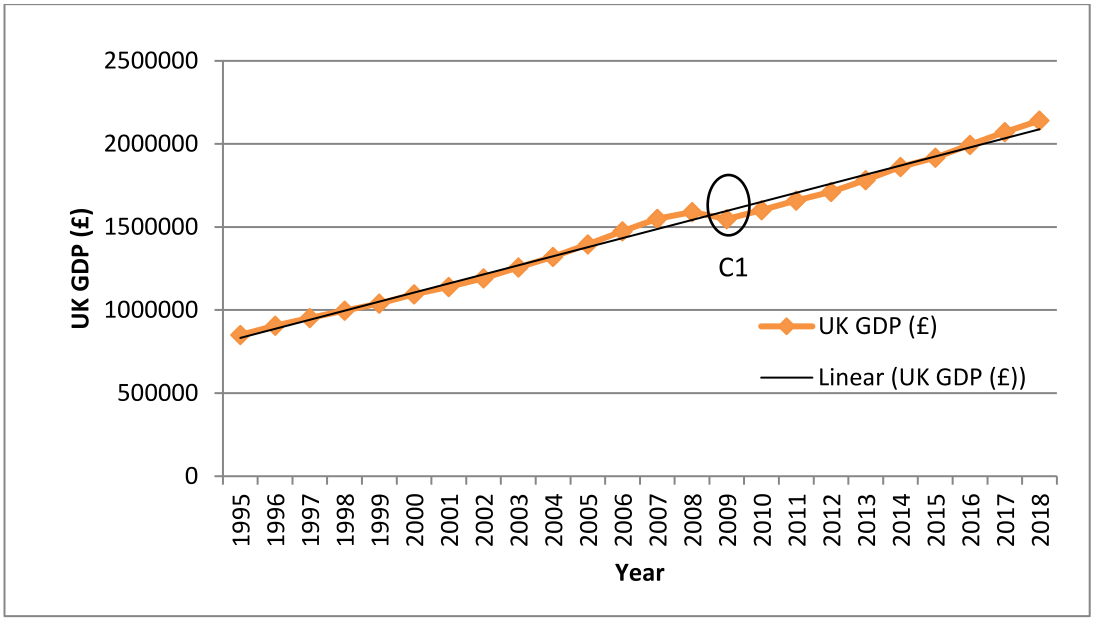


Figure 7 displayed the United Kingdom Gross Domestic Product (UK GDP) from 1995 until 2018 which comprises of 24 years. The GDP value decreased from 1.59 million in 2008 to 1.55 million in 2009 with a sudden dropped of 2.7%. This was due to global economic crisis but has affected UK GDP severely and considered as the great recession because it was a long term effect and took 5 years to recover (Velma & James, 2018). This drop in 2009 has shifted the trend line about 6% lower and the GDP only bounced back in 2014; but still growing slowly onwards to 2018 (Cribb & Johnson, 2018). Based on the UK GDP time series plot (Figure 6) and also the reasons behind those values, we can confirmed that the UK GDP data has trend and cyclical components. The trend was linear and the cyclical (C1) was due to the great recession (Figure 7).

Two sets of linear trend data have been used earlier (Ibadan Monthly rainfall and UK GDP) where BFTSC and GFTSC performed well in both data sets. Next, BFTSC and GFTSC were evaluated using another two data sets having quadratic curve trend (the London Stock Exchange (LSE) and the United States (US) Stock Market).

London Stock Exchange (LSE) Data

Figure 7, 8 and 9 shows the plots produced by BFTSC, GFTSC and time series plot for LSE monthly data respectively. Figure 7 is the time series plot for LSE monthly data using BFTSC. Figure 8 is the time series plot for LSE monthly data using GFTSC. Both BFTSC and GFTSC failed to identify curve trend and display linear trend instead. They also identify only one cyclical and no irregular which contradict with manual approach identification in figure 9 as in previously. These indicated the limitation of BFTSC and GFTSC when the trend deviated from linear which reflected similar findings as in the simulation study.

Figure 9 displays the manual time series plot of monthly LSE from January 2001 until December 2020. There was steady increment from 2001 until 2007 but dropped in 2008 due to economic crisis and slowly increased from 2009 to 2017. However, dropped once again in 2018 due to economic crisis. Then, it started to increase regularly up to 2019 but drastically dropped in 2020 due to COVID-19 pandemic (Moradi, Jabbari. & Rounaghi, 2021; Baker, Bloom, Davis, Kost, Sammon & Viratyosin, 2020). The first two dropped (in 2008 and 2018) related to economic crisis was considered as cyclical component (C1 and C2) and the third dropped (in 2020) related to COVID-19 pandemic was irregular component (I1). Figure 9 also shows a curve trend.

***Figure 8* . BFTSC plots of Monthly LSE**

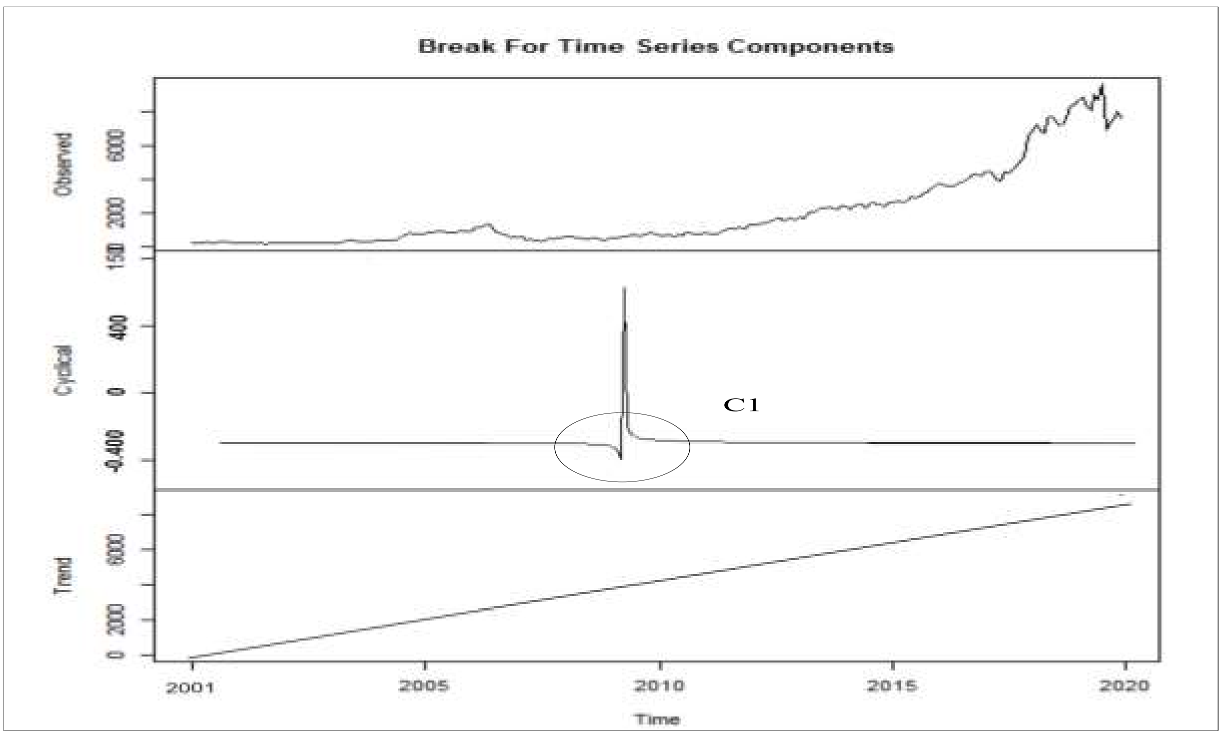


***Figure 9* . GFTSC plots of Monthly LSE**

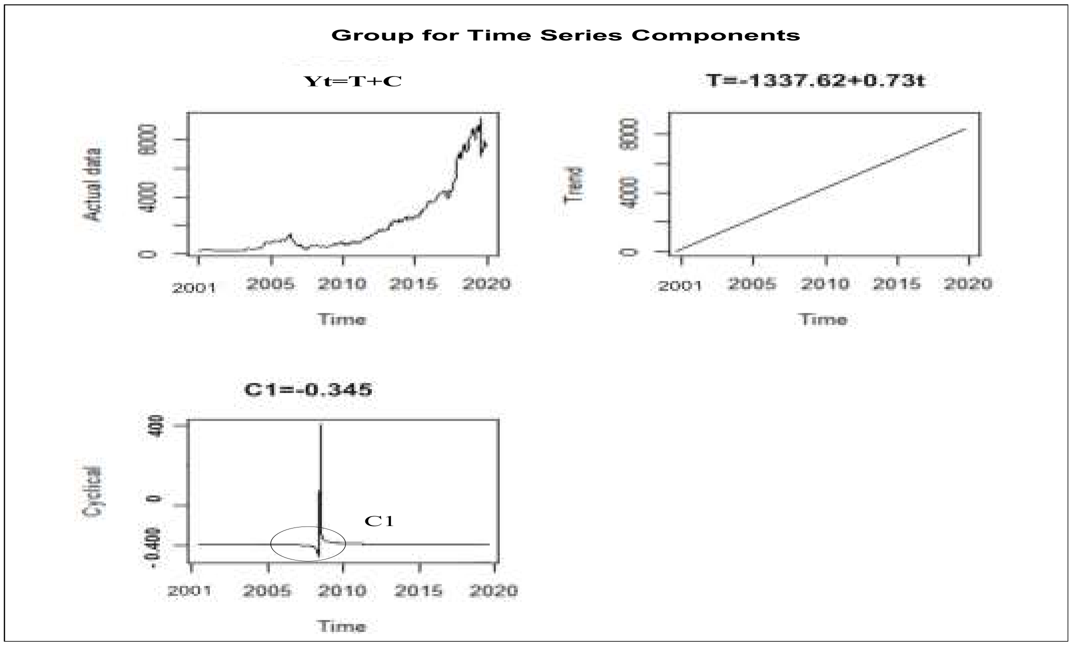
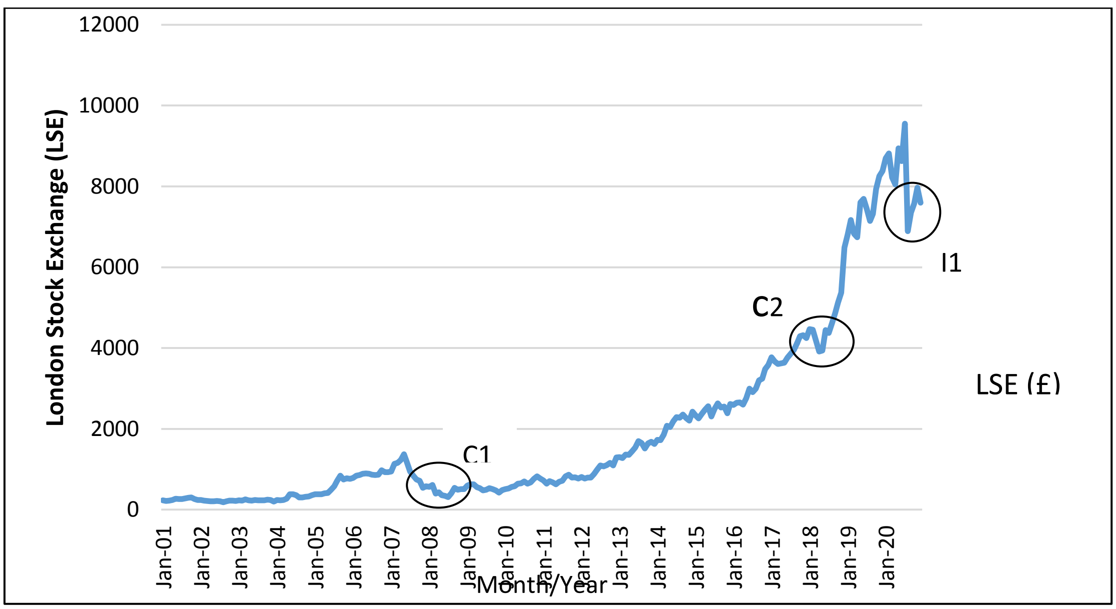


Figure 10. **TIME SERIES PLOT OF MONTHLY (LSE)**



United States (US) Stock Market Data

Figure 10, 11 and 12 show the plots produced by BFTSC, GFTSC and time series plot for US stock market monthly data respectively. Figure 10 is the time series plot for US stock market monthly data using BFTSC. Figure 11 is the time series plot for US stock market monthly data using GFTSC. BFTSC and GFTSC has not performed well for US Stock Market monthly data respectively. Both BFTSC and GFTSC managed to identify one cyclical but failed to identify curve trend and display linear trend instead, which contradict with manual approach identification as in previously. These indicated the limitation of BFTSC and GFTSC when the trend deviated from linear which reflected similar findings as in LSE data and the simulation study

Figure 12 is the time series plot of monthly US Stock Market. In this study, the US Stock Market was a monthly data from January 2001 until December 2018 and a total of 18 years. The data was obtained from the Yahoo Finance using Nasdaq adjusted close data which was the closing price after adjustment for all applicable splits and dividend distribution (Tang, Xiao, Wahab & Ma, 2021). The measurement of US Stock Market is in United States Dollars ($). There was steady increment from 2001 until 2007 but dropped in 2008 due to economic crisis and slowly increased from 2009 to 2018 (Shirvani, 2020). The dropped in 2008 that related to economic crisis was considered as cyclical component (C1). Figure 12 also shows a curve trend.

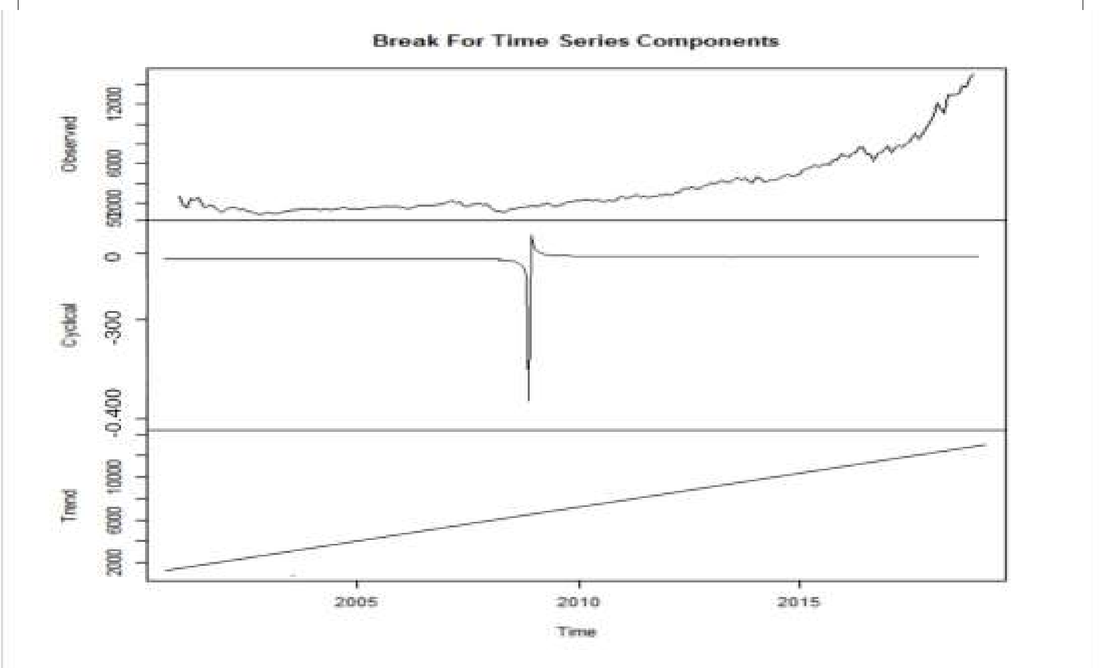


Figure 11 BFTSC plots Of monthly US stock market

Figure 12. GFTSC plots of Monthly US Stock Market

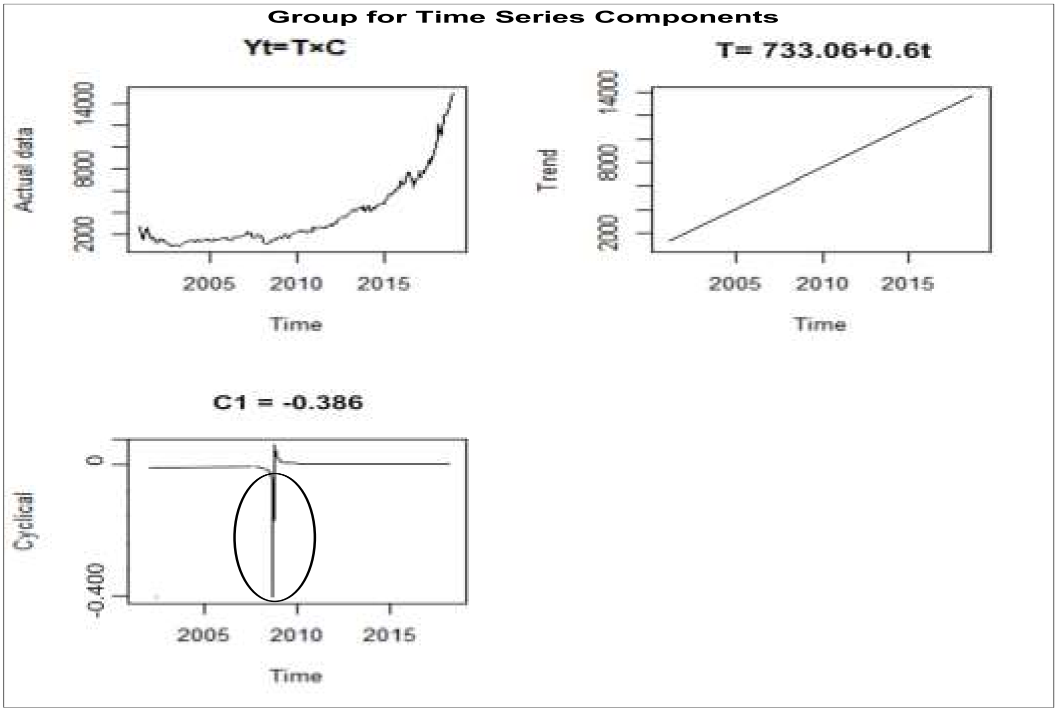
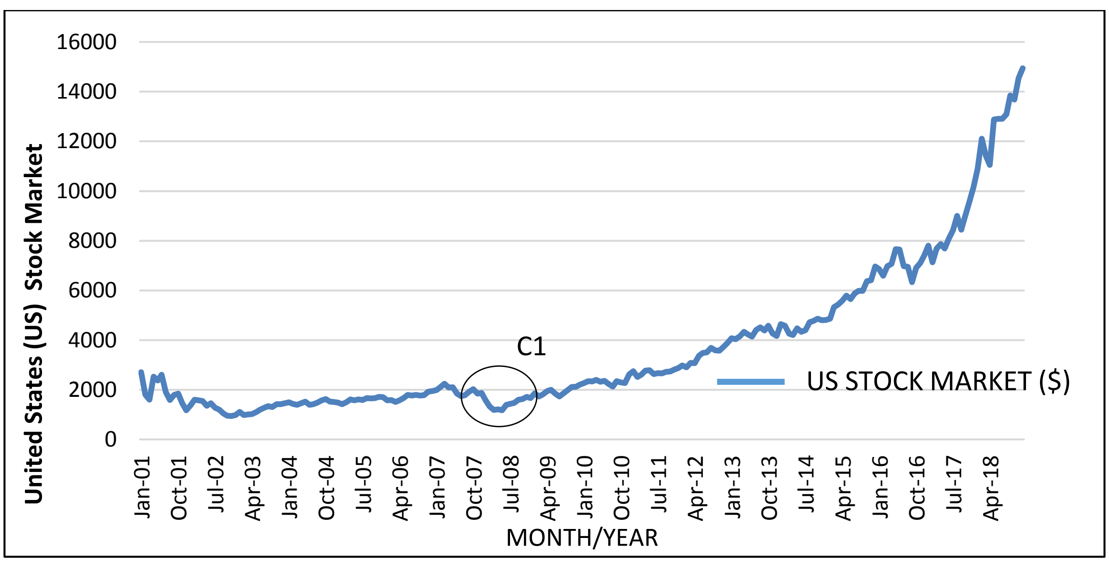


FIGURE 13 . Time Series plots of monthly US Stock Market



Overall BFTSC and GFTSC performed very well in identifying the time series components that embedded in the first two empirical data of Ibadan rainfall and UK GDP that showing linear trend. Thus, the displaying of the plots automatically and simultaneously makes time series identification process easy, straight forward and fast for end users. However, BFTSC and GFTSC performed poorly when using LSE and US Stock Market data that exhibiting curve trend. These indicated the limitation of BFTSC and GFTSC when the trend deviated from linear which reflected similar findings as in the simulation study. This is due to the derivation of BFTSC and GFTSC was based on linear trend as in BFAST. Thus, to use BFTSC and GFTSC requires large sample size to identify the existence of linear trend and other time series components (seasonal, irregular, and cyclical) in a data set.

Discussion

Box and Jenkins (1976) was one of the first researchers that struggle to clearly identify time series component using time plot. The first time series data obtained was plotted on a time plot using manual technique and the behavior of time series data was observed. However, the limitation of this technique was the complexity, it was very complicated to differentiate the time series components using casual manual time plot and the manual technique may be extremely difficult for non-experts. Ewing and Malik (2013) developed DBEST (Detection Breakpoint and Estimating Segment Trend) which was modified from BFAST. DBEST take in (NDVI) normalizes difference vegetation index data. The limitation of DBEST technique is that, the algorithm was built to solve the problem of topographical vegetation trend identification but cannot identify cyclical and irregular components of time series statistics. It is not flexible time series component identification technique and this is still the BFAST weakness that needs to be fully addressed. Jong, Verbesselt, Schaepman and Bruin (2012) argue and contributed to the body of knowledge by investigating the collective change identification called BFAST. The technique called BFAST is used for acknowledging breaks for additive, seasonal and trend in order to classify seasonal components and also enables the identification of breaks that take place in trend within the system (Morrison et al., 2018). Verbesselt, Zeileis, Hyndman, & Verbesselt (2012). Package 'bfast'which portrays the main scope of BFAST. Many scholars employ the use of BFAST in identifying trend in topographical data (Porter & Zhang, 2018). The technique is accessible in BFAST pack for R (R developments Core Team, 2012). Jain, Duin, and Mao (2000) describe time series component identification as a complicated issue, this lead this study to seek out for transparency regarding BFAST. Verbesselt et al. (2010) recommend a new technique for broad trend detection, image classification and representative, the technique is called Break for Additive Seasonal and Trend known as BFAST. This technique integrates the decomposition of time series components into the conventional elements of the series such as data, seasonal, trends and remnants, it was done with the help of the technique for identifying change which is embodied in the system of BFAST (Abbes & Farah, 2017; Adewoye & Chapman, 2018).

Therefore, from these discussion, BFAST was extended to a technique that can identify the four time series components. BFTSC is recommended for efficient time series components identification for an improved forecasting.

Conclusion

Based on the every result in the simulated and the empirical analysis, BFTSC and GFTSC is the most appropriate for time series components identification, for this reason BFTSC and GFTSC is recommended as a good alternative to BFAST. This is because BFTSC and GFTSC identifies the four components of time series statistics which is one of the basic limitations of BFAST. GFTSC also outperform BFTSC with 0.2%. Based on the forecast value for a years ahead, it reveal no evidence of water disaster in the period 2019 in Ilorin Kwara state.

BFTSC and GFTSC produced the related plots automatically and showing them simultaneously. There was a slight difference between BFTSC and GFTSC when displaying the plots where BFTSC combined the plots while GFTSC separated them and included equation and time series components values on top of the respective plots. Hence, provides better understanding regarding the time series components and can be used in forecasting. Overall, based on the simulation and empirical findings, we can conclude that both BFTSC and GFTSC are performing very well in large data set that displaying linear trend which can bridge the gap between expert and end users in identifying the time series components because they are easy, straight forward and fast. These findings indicated that BFTSC and GFTSC automatic identification techniques are suitable for data with linear trend and require future extensions for other trends.

Ethics

This is the original manuscript; there will be no expectation of any ethical problems after the publication. The two authors have read and approved the manuscript.

References

Abbes, A. B., & Farah, I. R. (2017). Prediction Changes for Nonstationary Multi- Temporal Satellite Images using HMM. In *Handbook of Research on Geographic Information Systems Applications and Advancements* (pp. 387-406). IGI Global. DOI: 10.4018/978-1-5225-0937-0.ch015

Ajare, Emmanuel Oloruntoba , Suzilah bt. Ismail, (2019). Simulation of Data to Contain the Four Time Series Components in Univariate Forecasting. *Journal of Advanced Research in Dynamical and Control Systems*, 11(5), 1005 - 1010.

Ajare, Emmanuel Oloruntoba , Suzilah bt. Ismail, (2019). Break for Time Series Components (BFTSC) and Group for Time Series Components (GFTSC) in Identification of Time Series Components in Univariate Forecasting. *Journal of Advanced Research in Dynamical & Control Systems*, 11(5), 995 - 1004.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons. Retrieve from www.scirp.org/

Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control, revised ed*. Holden-Day. Oakland, California. Retrieve from garfield.library.upenn.edu/classics

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics*, *18*(1), 1-22. **https://doi.org/10.1002/jae.659**

Bornhorst, F., Dobrescu, G., Fedelino, A., Gottschalk, J., & Nakata, T. (2011). When and how to adjust beyond the business cycle? A guide to structural fiscal balances. *IMF Technical Notes and Manuals*, *11*(02). Retrieve from https://www.imf.org/external/

Bohn, H. (1995). The sustainability of budget deficits in a stochastic economy. *Journal of Money, Credit and Banking*, *27*(1), 257-271. DOI: 10.2307/2077862 https://www.jstor.org/stable/2077862

Buhalau, T. (2016). Detecting clear-cut deforestation using Landsat data: A time series a nalysis of remote sensing data in Covasna County, Romania between 2005 and 2015. *Student thesis series INES*. Retrieve from https://lup.lub.lu.se/student papers/search/publication/

Cesta, A., Cortellessa, G., Pecora, F., & Rasconi, R. (2005, May). Monitoring Domestic Activities with Scheduling Techniques. In *Proceedings of the 2nd*. Retrieve from https://www.researchgate.net/publication/238572582

Cleveland, W. P., & Tiao, G. C. (1976). Decomposition of seasonal time series: A model for the Census X-11 program. *Journal of the American statistical Association*, *71*(355), 581- 587. Retrieve from https://amstat.tandfonline.com/doi/abs/

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for information Science and Technology*, *57*(3), 359-377. **https://doi.org/10.1002/asi.20317**

Caiado, J. (2009). Performance of combined double seasonal univariate time series models for forecasting water demand. *Journal of Hydrologic Engineering*, *15*(3), 215-222.

Cipra, T., & Romera, R. (1997). Kalman filter with outliers and missing observations. *Test*, *6*(2), 379-395. Retrieve from https://link.springer.com/article/10.1007/BF02564705

DeVries, B., Pratihast, A. K., Verbesselt, J., Kooistra, L., de Bruin, S., & Herold, M. (2013, June). Near real-time tropical forest disturbance monitoring using Landsat time series and local expert monitoring data. In *Analysis of Multi-temporal Remote Sensing Images, MultiTemp 2013: 7th International Workshop on the* (pp. 1-4). IEEE **DOI:** 10.1109/Multi-Temp.2013.6866022

Ewing, B. T., & Malik, F. (2013). Volatility transmission between gold and oil futures under structural breaks. *International Review of Economics & Finance*, *25*, 113-121.journal homepage: Retrieve from www.elsevier.com/locate/iref

Flicek, P., & Birney, E. (2009). Sense from sequence reads: methods for alignment and assembly. *Nature methods*, *6*(11s), S6. Retrieve from https://www.ncbi.nlm.nih.gov/pubmed/19844229

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary- scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18-27 . https://doi.org/10.1016/j.rse.2017.06.031

Jong, R., Verbesselt, J., Schaepman, M. E., & Bruin, S. (2012). Trend changes in global greening and browning: contribution of short-term trends to longer-term change. *Global Change Biology*, *18*(2), 642-655. DOI:10.1111/j.1365- 2486.2011.02578.x

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, *22*(1), 4-37. **DOI:** 10.1109/34.824819

Jamali, S., Jönsson, P., Eklundh, L., Ardö, J., & Seaquist, J. (2015). Detecting changes in vegetation trends using time series segmentation. *Remote Sensing of Environment*, *156*, 182-195. doi.org/10.1016/j.rse.2014.09.010

Porter, J., & Zhang, L. (2018). BisPin and BFAST-Gap: Mapping bisulfite-treated reads. *bioRxiv*, 284596. **doi:** https://doi.org/10.1101/28459

Tolsheden, J. (2018). Detecting and testing for Seasonal breaks in Quarterly National accounts: Based on X-12-ARIMA and BFAST Methods. Retrieve from diva-portal.org

Maggi, L. M. B. (2018). Times Series Analysis. In *Multiscale Forecasting Models* (pp. 1-29). Springer, Retrieve from https://www.springer.com/us/book/9783319949918 Mok, T. S., Wu, Y. L., Ahn, M. J., Garassino, M. C., Kim, H. R., Ramalingam, S. S., ... & Lee, C. K. (2017). Osimertinib or platinum–pemetrexed in EGFR T790M–positive lung cancer. *New England Journal of Medicine*, *376*(7), 629-640. DOI: 10.1056/NEJMoa1612674

Maus, V., Câmara, G., Appel, M., & Pebesma, E. (2017). dtwSat: Time-Weighted Dynamic Time Warping for Satellite Image Time Series Analysis in R. *Journal of Statistical Software*. Retrieve from https://cran.r-project.org/web/packages/dtwSat/ind...

Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. *Remote sensing of Environment*, *114*(1), 106-115. doi.org/10.1016/j.rse.2009.08.014

Rikus, L. (2018). A simple climatology of westerly jet streams in global reanalysis datasets part 1: mid-latitude upper tropospheric jets. *Climate Dynamics*, *50*(7-8), 2285-2310.

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint*. Retrieve from https://arxiv.org/pdf/1703.10593

Zewdie, W., Csaplovics, E., & Inostroza, L. (2017). Monitoring ecosystem dynamics in northwestern Ethiopia using NDVI and climate variables to assess long term trends in dry land vegetation variability. *Applied Geography*, *79*, 167-178. https://doi.org/10.1016/j.apgeog.2016.12.019

Zdravevski, E., Lameski, P., Mingov, R., Kulakov, A., & Gjorgjevikj, D. (2015, September). Robust histogram-based feature engineering of time series data. In *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on* (pp. 381-388). IEEE. **DOI:** 10.15439/2015F420

Zhao, G., Li, E., Mu, X., Wen, Z., Rayburg, S., & Tian, P. (2015). Changing trends and regime shift of streamflow in the Yellow River basin. *Stochastic environmental research and risk assessment*, *29*(5), 1331-1343. Retrieve from https://www.hindawi.com/journals/amete/2016/6130179/ref/

Zeileis, A., Kleiber, C., Krämer, W., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis*, *44*(1-2), 109-123. https://doi.org/10.1016/S0167-9473(03)00030-6