Jurnal
Teknologi

Full Paper

PERFORMANCE OF 4253HT SMOOTHER BY DIFFERENT HANNINGS: APPLICATION IN RAINFALL DATA

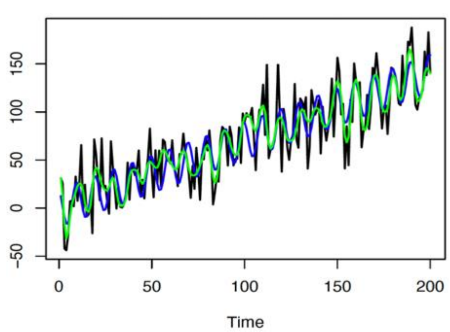
Adie Safian Ton Mohamed[a*], Noor Izyan Mohamad Adnan[b], Qasim Nasir Husain[c], Adina Najwa Kamarudin[d], Nurul Nisa' Khairol Azmi[e]

[a]School of Mathematics, Actuarial and Quantitative Studies, Asia Pacific University of Technology and Innovation, 57000 Kuala Lumpur, Malaysia
[b]College of Computing, Informatics and Media, University Teknologi MARA Pahang, Jengka Campus, 26400 Jengka, Pahang, Malaysia
[c]Department of Mathematics, College of Education for Pure Science, Tikrit University, Iraq
[d]Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
[e]College of Computing, Informatics and Media, University Teknologi MARA Negeri Sembilan, Seremban Campus, 70300 Seremban, Negeri Sembilan, Malaysia

Article history
Received
4 January 2023
Received in revised form
6 July 2023
Accepted
12 July 2023
Published Online
20 October 2023

*Corresponding author
adie.safian@staffemail.apu.edu.my

Graphical abstract



Abstract

Smoothing is an exploratory data analysis approach that focuses on removing noise or unstructured pattern from data series. This study mainly aims to compare the performance of 4253HT smoother in three types of Hannings and its application in forecasting. A sinusoidal signal was used where five different levels of contaminated normal noise were applied. Overall, 4253HT smoother with Shitan and Vazifean's Hanning performs excellently over different percentages of noise, good at preserving edges, and able to travel closely with the signal of original pattern. The smoothed rainfall data gives a lower value of RMSE than the raw data which is 12.85 and 24.25 respectively. This concludes that the trend line obtained using smoothed data is more appropriate and reliable for forecasting. These results will be useful in predicting any time series data.

Keywords: Nonlinear smoother, 4253HT, extreme data, rainfall, noise

Abstrak

Pelicinan merupakan sebuah kaedah dalam penerokaan analisa data yang memfokuskan kepada penyingkiran hingar dari sebuah siri data. Kajian ini bertujuan untuk membandingkan prestasi pelicin 4253HT yang menggunakan tiga jenis Hanning dan aplikasinya dalam kaedah peramalan. Isyarat sinusoidal telah digunakan dimana lima tahap gangguan telah diaplikasikan. Secara keseluruhannya, pelicin 4253HT yang digunakan bersama Hanning Shitan dan Vazifean menunjukkan hasil yang cemerlang dalam peratusan hingar yang berbeza, dapat mengekalkan keseimbangan data dengan baik, dan mampu bergerak selari dengan isyarat yang bercorak asal. Data yang licin memberi nilai

RMSE yang lebih rendah berbanding data asal, iaitu dengan nilai masing-masing 12.85 dan 24.25. Ini dapat disimpulkan bahawa garis arah aliran yang diperoleh menggunakan data yang licin adalah lebih sesuai dan lebih dipercayai untuk ramalan terhadap data bersiri masa.

Kata kunci: Pelicin tidak lurus, 4253HT, data ekstrim, taburan hujan, hingar

1.0 INTRODUCTION

Smoothing is an exploratory data analysis approach that focuses to remove noise or unstructured patterns from data series, making the patterns or signals more visible. There are two types of data smoother, which are conventional linear smoother and non-linear smoother. Linear smoothers are the most commonly used and work excellently with Gaussian noise [1]. However, the efficacy of linear smoother is limited in certain conditions due to the nature in data smoothing such as a destruction by noise from heavy-tailed and unknown distribution [2]. To overcome this limitation, the non-linear smoother is introduced as it has been proven to be more robust to heavy-tailed noise than the linear smoother based on previous studies [3, 4, 5, 6].

Non-linear smoother such as running median was firstly introduced by Tukey (1977) and his works, later was extensively expanded by Rabiner *et al.,* (1975). These non-linear smoothers are widely used in engineering practices, including image and speech processing applications [5, 9]. Moreover, this method also was applied in signal processing by Masrelies and Martin (1977), and Azmi *et al., (2018)*. Over the decades, Tukey's and Vellemen's idea had resulted to a few novel smoothers such as 53H (or 53HT), 3RSSH (or 3RSSHT), and 4253H (or 4253HT) where all these smoothers are currently known as compound smoothers.

Compound smoother is a smoothing method that mixes several different smoothing methods, including median smoother of various span windows, hanning or weighted moving average, splitting and re-smoothing the residual [12]. For instance, smoother (4253H, twice) was initially introduced by Jin and Xu (2013) which involved the running median of even and odd window size, hanning and 'twice'. Studies by Shitan and Vazifedan (2011), and Velleman (1980) came out with a new combination of median smoother for a better smoothing result. Shitan and Vazifedan (2011), and Velleman (1980) found that forecast using smoothed data gives more advantages in forecasting compared to raw data and also acknowledged that compound smoother, 4253HT is the best non-linear smoother [15, 18]. The latest hanning type was introduced by Husain (2017) with different combination of coefficients were used. Hence, based on the past research, there are several types of hanning which will affect the performance of smoothing.

Since hanning is one of the important factors in observing smoother's performance, a study on comparison of hanning types should be conducted. Moreover, studies on the comparison of hanning's performance in smoothing technique has not been conducted extensively by scholar previously. There are great potentials and opportunities to conduct modifications of hanning. Throughout this study, the performance of 4253HT smoother with three different types of hanning will be demonstrated in sinusoidal signal recovery, where five different levels of contaminated normal noise (10%, 25%, 50%, 75%, and 90%) will be applied. The three different types of hanning which will be used are Tukey (1977), Shitan and Vazifean (2011), and Husain (2017). Hanning with the greatest potential to performs over different percentages of noise, preserve edges, and travel closely with the signal of original pattern will be observed and determined. Afterwards, the best hanning type will be used in forecasting of a rainfall data.

2.0 METHODOLOGY

The focus in this section is on the performance of 4253HT smoother using three types of hanning in signal recovery from contaminated normal noise by using simulations. A sinusoidal signal with the addition of contaminated normal noise were used in this study. The performance was measured using estimated integrated root mean square error (EIRMSE). Different hanning coefficient proposed by Tukey (1977), Shitan and Vazifean (2011), and Husain (2017), were applied to 4253HT, then was discussed, and compared. The main R programming algorithm of 4253HT is as described in appendix, and the details is as designated by Velleman (1980).

2.1 4253HT Smoother

A compound smoother 4253HT is a combination of various algorithm that starts with a running median of window four, followed by running median of two for re-center purpose. The result was re-smoothed by a running median of five and another running median of three. Next, hanning, denoted by H was applied. Then final step called "Twicing" was performed, where the same smoother was applied to the residual, and added to first pass of smoother.

Besides, even-window medians were introduced in this algorithm to alleviate some of the problems associated with odd-window medians. Moreover, the function of twicing is to recover patterns from the residuals.

Sequence **x** is described as an infinite series of phenomenon which is written as $x_{i-n}, \dots, x_{i-1}, x_i, x_{i+1}, \dots x_{i+n}$, where x_i denotes an observed series at i^{th} time. Meanwhile, smoother *S* refers to an algorithm that applies **x** to obtain a new series of smoothed values $S(x_i)$. The 4253HT smoother is applied according to the following steps:

Step 1: Conduct running median of window four and re-centered by running median of window two:

$$S_4(x_i) = median\ [x_{i-2}, x_{i-1}, x_i, x_{i+1}] \quad (1)$$
$$S_{42}(x_i) = median\ [S_4(x_i),\ S_4(x_{i+1})] \quad (2)$$

Step 2: Conduct running median of window size five and three:

$$S_{425}(x_i) = median[S_{42}(x_{i-2}),\ S_{42}(x_{i-1}),\ S_{42}(x_i), S_{42}(x_{i+1}), S_{42}(x_{i+2})] \quad (3)$$
$$S_{4253}(x_i) = median[S_{425}(x_{i-1}),\ S_{425}(x_i), S_{425}(x_{i+1})] \quad (4)$$

Step 3: Perform hanning, H using coefficients $\left\{\frac{1}{4},\frac{1}{2},\frac{1}{4}\right\}$ as an illustration for algorithm, which is as follows:

$$S_{4253H}(x_i) = \frac{1}{4}S_{4253}(x_{i-1}) + \frac{1}{2}S_{4253}(x_i) + \frac{1}{4}S_{4253}(x_{i+1}) \quad (5)$$

Step 4: Smooth the residual and add it back to the smoothed value:

$$e_i = x_i - S_{4253H}(x_i) \quad (6)$$
$$S_{4253HT}(x_i) = S_{4253H}(x_i) + S_{4253H}(e_i) \quad (7)$$

"Hanning", the name after an Austrian meteorologist Julius von Hann, also known as weighted moving average or mean. It was suggested by Velleman and Hoaglin (1981) as one of smoothing techniques.

It can be classified into equal and unequal weighted averages. The most common mathematical hanning form is the unequal weighted method introduced by Tukey (1977), which later was used by Velleman and Hoaglin (1981). Meanwhile, the equal weighted method was suggested by Husain (2017).

List of hanning coefficient that were used in this study are as follows:

Tukey (1977):

$$A = \left\{\frac{1}{4},\frac{1}{2},\frac{1}{4}\right\} \quad (8)$$
$$H_i = \frac{1}{4}(x_{i-1}) + \frac{1}{2}(x_i) + \frac{1}{4}(x_{i+1}) \quad (9)$$

Shitan and Vazifean (2011):

$$A = \left\{\frac{1}{3},\frac{1}{3},\frac{1}{3}\right\} \quad (10)$$
$$H_i = \frac{1}{3}(x_{i-1}) + \frac{1}{3}(x_i) + \frac{1}{3}(x_{i+1}) \quad (11)$$

Husain (2017):

$$A = \left\{\frac{3}{8},\frac{2}{8},\frac{3}{8}\right\} \quad (12)$$
$$H_i = \frac{3}{8}(x_{i-1}) + \frac{2}{8}(x_i) + \frac{3}{8}(x_{i+1}) \quad (13)$$

2.2 Recovery of Signal

Generally, the data, x_i can be expressed as the sum of two components:

$$X_i = S_i^* + W_i \quad (14)$$

with S_i^* denoting the signal and W_i is the noise. Recovery of signal is the process by which a noise is eliminated from a data to obtain a signal. The signal that was contaminated by normal noise was used during the simulation process to determine the performance of 4253HT smoother. All the simulation steps were done by referring to Conradie *et al.,* (2009).

2.2.1 Simulation Procedure

A signal, S_i^* used in this simulation was a sinusoidal function with linear trend:

$$S_i^* = \alpha\sin\beta(i-\phi) + D_i \quad (15)$$

The parameters for this sinusoidal function are denoted as; α is amplitude, D is linear component, i is index, $\beta = 2\frac{\pi}{p}$ with $\frac{1}{p}$ is frequency and p is period, then ϕ is displacement. This is the similar signal used by Conradie *et al., (2009)* and Coles (2001) with the respective parameter values to yield a smooth and good curve: $|\alpha| = 3$, $\phi = 1$, and $D = 0.7$.

The noise, W_i was generated as independent and identically distributed random variables from a contaminated of two normal distributions. The selected parameters used to generate a noise with higher volatility are $W_{1i}{\sim}N(0,1)$ and $W_{2i}{\sim}N(0, 5.06^2)$ [12]. Since the interest was particularly in data with high kurtosis, the variance of W_{2i} was taken as 5.06^2. There were five levels of contaminated normal noise; 10%, 25%, 50%, 75%, and 90%. The high levels of contamination were considered in observing the performance of smoother when a data is corrupted with extreme noises. The simulation of 10% contaminated normal distribution indicates that approximately 10% of the values come from a $N(0, 5.06^2)$ distribution and approximately 90% from a $N(0,1)$ distribution [19].

2.3 Real Data

Based on the literature review, the old method of forecasting was performed by using the whole retrieved data. However, the use of extracted signal as a backbone of dataset for forecasting is applied in this study. A daily rainfall (millimeter) data retrieved from Stesen Keretapi Kerambit in Pahang is the only data that has missing values of less than 10%. However, the daily rainfall data was not able to be used in this study as the 4253HT smoother is not robust with a dataset that has missing values. As a solution, Block Maxima which is an extreme value data approach was performed [19]. This study focuses on the maximum values as a data, where only monthly maximum rainfall from January 1987 to December 2016 were considered.

Pahang is such a huge state as it is the third largest region of Malaysia after Sabah and Sarawak. The location of Pahang is as illustrated in Figure 1, which is represented in red. Besides, Pahang is geographically varied, consists of massive Pahang River basin that is bounded with Titiwangsa Mountains extended towards the west, and the eastern highlands towards the north. Although huge part of Pahang is occupied with dense rainforest, but Pahang River leads the drainage system as it is formed by several rivers which are interconnected at the central plains. Moreover, Pahang receives maximum rainfall from November until January, while minimum rainfall occurs in June and July.

Figure 1 Location of Pahang

2.4 Forecasting

The method of Holt-Winters Exponential Smoothing was performed in this study to forecast time series data that exhibits both trend and seasonal variations [20].

The efficacy of using raw data or smoothed data (smoothed by 4253HT) in forecasting was examined to determine which method gives the best trend line in representing the data. The performance was measured by using Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{j=1}^{m}(X_j - \hat{X}_j)^2}{m}} \tag{16}$$

Where X_i is observation (raw data or smoothed data) and X_i hat is trend value, then the constant, m is length of dataset.

3.0 RESULTS AND DISCUSSION

3.1 4253HT Performance Via Simulation

Figure 2 depicts a signal of sinusoidal with high frequency plus linear trend. When noise is added to the signal, it tends to distort the actual pattern and may lead to inaccuracy in deciding of appropriate method for forecasting. Figures 3 to 6 show sinusoidal signals with various levels of contaminated noise added.

The signal is can still be noticed with noise 25% and 50% applied. As the percentage of noise increases to 75% and 90%, the density of noise with spike increases, creating an interference to the signal, and making it difficult to detect the signal with naked eye. An increasing trend is spotted clearly in the figures. However, the wavelength of high frequency might be left out due to the noise that causes high volatility.

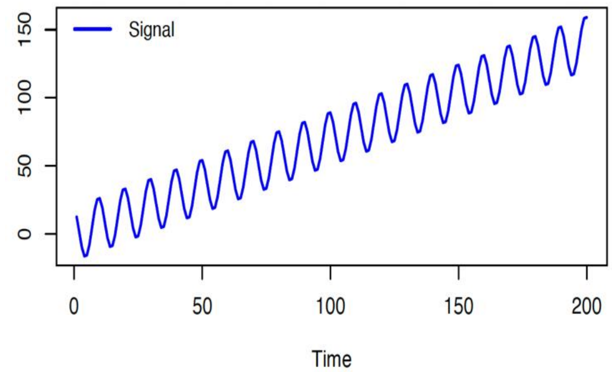


Figure 2 Linear sinusoidal function with a linear trend

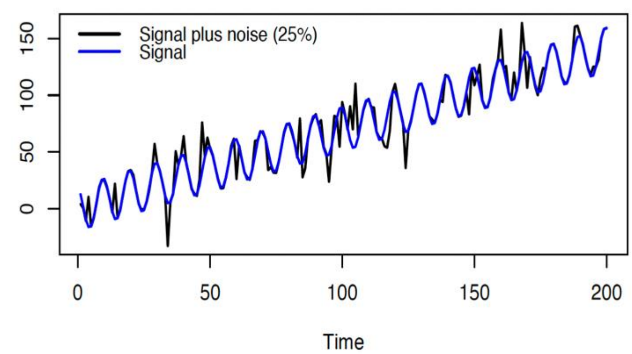


Figure 3 Linear sinusoidal function with 25% of contaminated normal noise

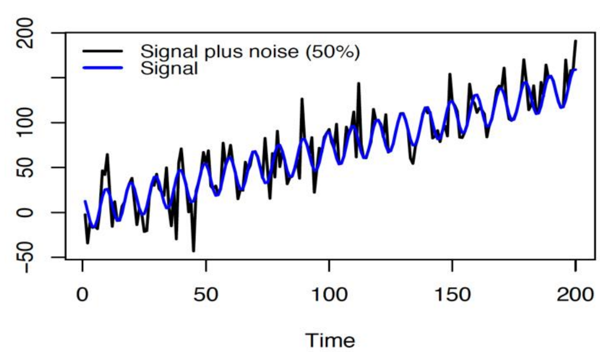


Figure 4 Linear sinusoidal function with 50% of contaminated normal noise

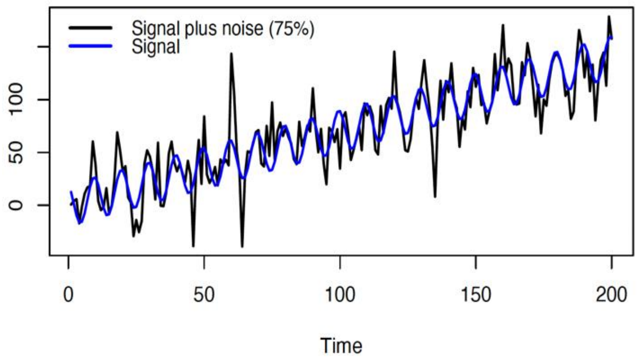


Figure 5 Linear sinusoidal function with 75% of contaminated normal noise

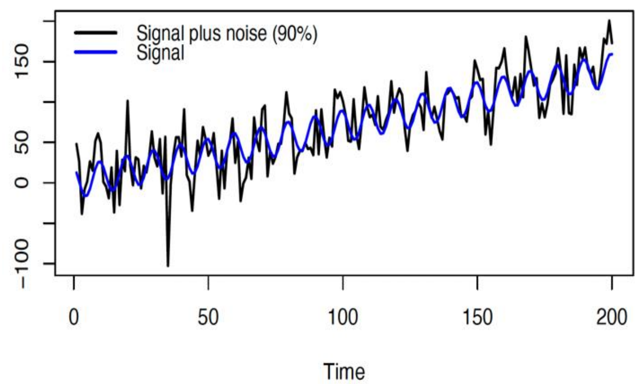


Figure 6 Linear sinusoidal function with 90% of contaminated normal noise

The simulation was run for 200 times, similar as the procedure conducted by Conradie *et al., (2009)*. The performance of 4253HT smoother was assessed by using estimated integrated root mean square error (EIRMSE) as an evaluator, which is written as:

$$EIRMSE = \sqrt{\frac{1}{k}\sum_{j=1}^{k}\frac{1}{n}\sum_{i=1}^{n}(S_{ij} - Z_j)^2} \tag{17}$$

where S_{ij} is an original noise-free signal, Z_j is the 4253HT smoother, and the constants *n* and *k* represent data length and number of simulations respectively. A low EIRMSE value indicates that a smoother performs well to eliminate normal noise.

Figures 7 to 11 depict that 4253HT smoother using Tukey (1977) hanning (green line) travels closely with signal (blue line) along the time, signifying that 4253HT smoother is preferable.

The best performance of 4253HT (highly capable) is observed when the lower noises (10% and 25%) were applied. Moreover, the smoother is still able to travel along with the blue trail when higher noises up to 90% (heavy noise) were applied. This indicates that it performs well even with a high percentage of noise. As the percentage of noise increases, the signal fluctuates unsteadily causing more difficult for the 4253HT to detect signal. This is supported by EIRMSE values provided in Table 1. It also shows that 4253HT is able to eliminate spike (see figures 9 and 10) and maintain its performance to detect smooth trails.

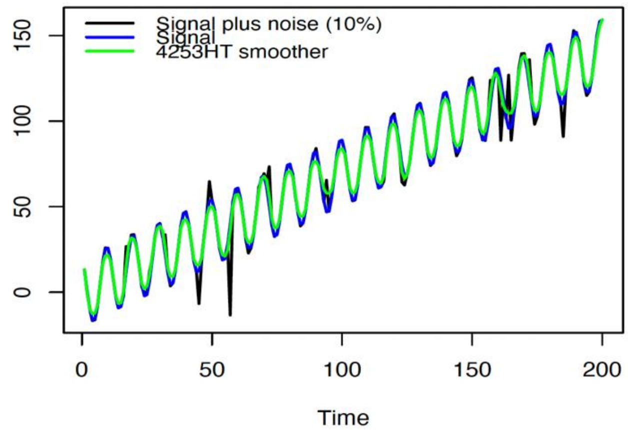


Figure 7 Performance of 4253HT in capturing signal with 10% of contaminated normal noise added

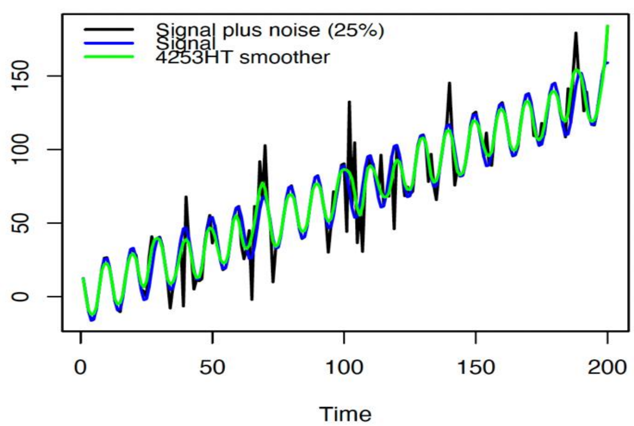


Figure 8 Performance of 4253HT in capturing signal with 25% of contaminated normal noise added

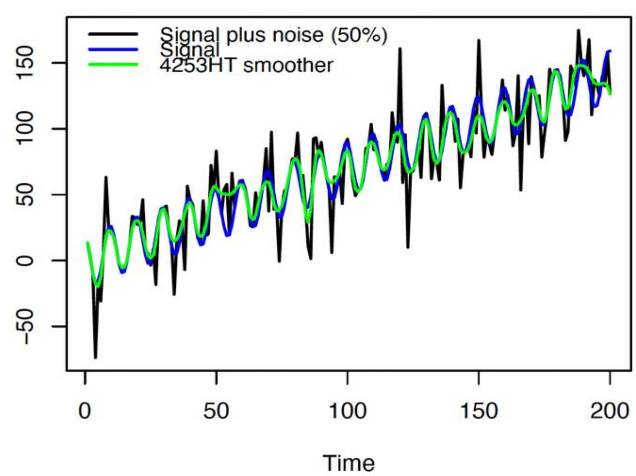


Figure 9 Performance of 4253HT in capturing signal with 50% of contaminated normal noise added

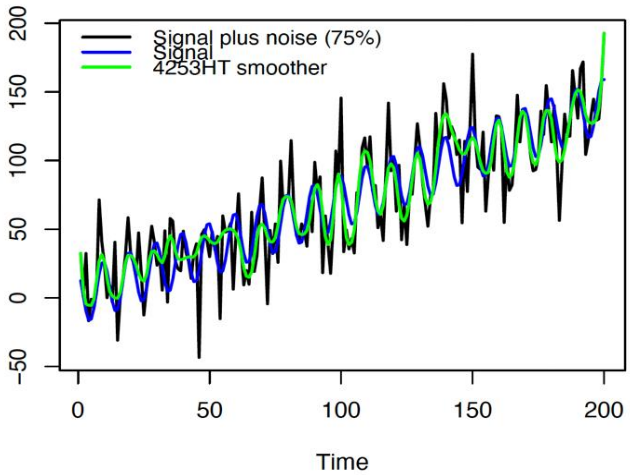


Figure 10 Performance of 4253HT in capturing signal with 75% of contaminated normal noise added

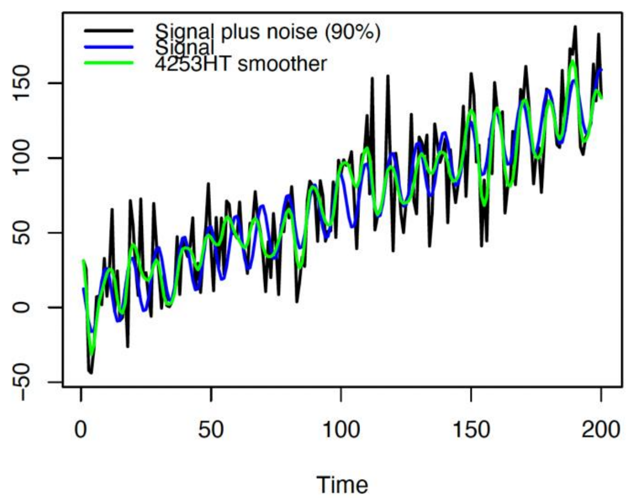


Figure 11 Performance of 4253HT in capturing signal with 90% of contaminated normal noise added

Overall, 4253HT performs excellently over different percentages of noise, able to preserve the edge well, and bears to travel closely with the signal original pattern.

Table 1 shows the numerical results (EIRMSE) for 4253HT with three types of hanning coefficient in extracting signal from the noise with contaminated normal distribution. The details for each smoother are as follows:

SM1: 4253HT smoother with hanning coefficient proposed by Tukey (1977)

SM2: 4253HT smoother with hanning coefficient proposed by Shitan and Vazifean (2011)

SM3: 4253HT smoother with hanning coefficient proposed by Husain (2017)

Generally, as the percentage of noise increases, the EIRMSE value for all smoothers increases with SM2 smoother performs better at lower noises (10% and 25%), while SM3 is preferred for higher noise (90%). This finding is supported by previous studies [3,4,5,6], where they claimed that the non-linear smoother is more robust to heavy-tailed noise than the linear smoother.

The result found that EIRMSE values are quite consistent and do not have great differences for all smoothers. Both SM2 & SM3 have advantages, but SM2 is able to perform over three levels of contaminated noise, which are 10%, 25%, and 75% compared to SM3 that is only able to perform over two levels. This concludes SM2 is the best hanning amongst them. Hence, hanning Shitan and Vazifean (2011) was chosen to perform the next analysis, which is forecasting.

Table 1 EIRMSE value to measure the performance of 4253HT in recovery sinusoidal signal from contaminated normal noise

	Contaminated Normal Noise (%)				
	10	**25**	**50**	**75**	**90**
SM1	3.6236	5.3208	8.1857	11.0923	12.6594
SM2	**3.6092**	**5.2987**	8.2777	**10.8585**	12.4986
SM3	3.7005	5.3654	**8.0992**	10.9256	**12.3922**

3.2 Application of 4253HT In Forecasting

This section will demonstrate the application of hanning that has the best performance, which is Shitan and Vazifean (2011) in the 4253HT smoother to perform forecasting in maximum rainfall data. The RMSE, Sigma and Akaike Information Criterion (AIC) values for raw and smoothed data will be compared to determine a suitable data that will give the most appropriate result in forecasting.

3.2.1 Trend Line

Figure 12 shows maximum amount of rainfall recorded by month with the trend line. Generally, it is very difficult to determine specific pattern that exists in the rainfall data due to outliers and high volatility. The application of smoothing 4253HT algorithm to the data leads to removal of some outliers by maintaining the main feature of the data series (seasonal fluctuation) as depicted in Figure 13. This allows us to easily suggest several methods that are appropriate and suit such behavior. Since the component of seasonality is observed, Holt's Winters methodology is the most recommended approach.

By using smoothed data, a clearer pattern of trend line can be obtained to observe the behaviors and pattern of data.

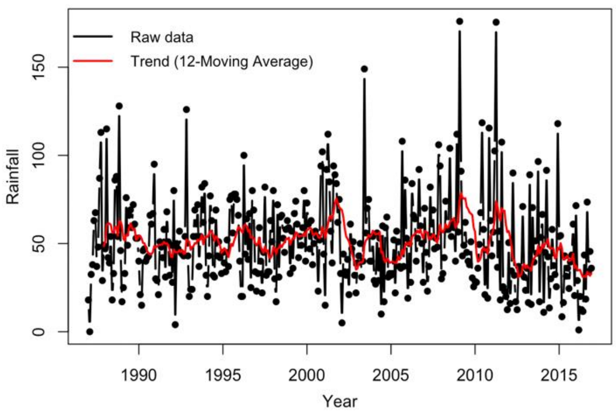


Figure 12 Scatter plot of raw data (rainfall) and trend line

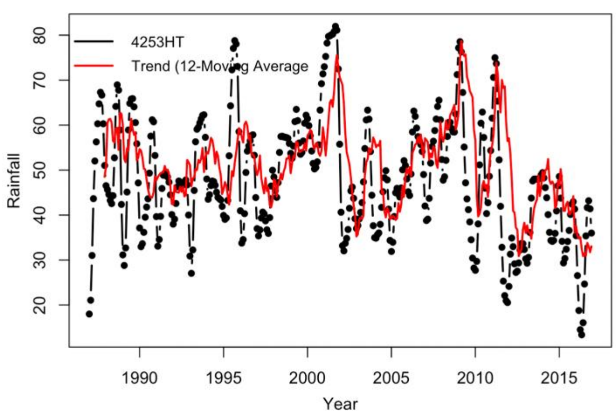


Figure 13 Scatter plot of smoothed data (4253HT) and trend line

Table 2 shows RMSE values for the distance between the trend and observation (raw data or smoothed data). The smoothed data gives a lower value of RMSE than the raw data. This indicates that the trend line created using smoothed data is more fit and more reliable for forecasting.

Table 2 RMSE values for the distance between the trend and observation (raw and smoothed data)

Data	RMSE
Raw	24.25248
Smoothed	**12.85097**

3.2.3 Holt Winter Forecasting

Both observation sets (raw and smoothed data) were divided into two parts, namely estimation and evaluation. In the first part of each set, 75% (270 observations) were used for the estimation of parameters (alpha, beta, and gamma). Whereas the rest 25% (90 observations) were used for evaluation to determine whether forecasting using smoothed data is better than using the actual data (raw data). The forecast performance indicator such as Sigma, Akaike Information Criterion (AIC), and Mean Absolute Percentage Error (MAPE) values were referred for a comparison purpose.

Table 3 shows the values of sigma, AIC and MAPE for forecasting using smoothed value are 4.1756, 2132.1590, and 21.5739 respectively. These values are lower than the values obtained for forecasting using raw data. The results signify that forecasting using smoothed values will produce a better forecast than using the raw data. The findings signify that smoothed values will generate a better forecast than using raw data is in line with [14,15], which found that the application of smoothed data provides more advantages in forecasting compared to raw data.

Table 3 Parameter values for raw and smoothed data

Data	Sigma	AIC	MAPE
Raw	22.3406	2977.4450	30.0900
Smoothed	**4.1756**	**2132.1590**	**21.5739**

Next, the smoothed data was used for forecasting of maximum rainfall for 12 periods (January to December 2017) by using Holt-Winters method. Figure 14 shows the smoothed data (4253HT) and fitted line (Holt Winter method). Based on this figure, the line fits the smoothed values perfectly.

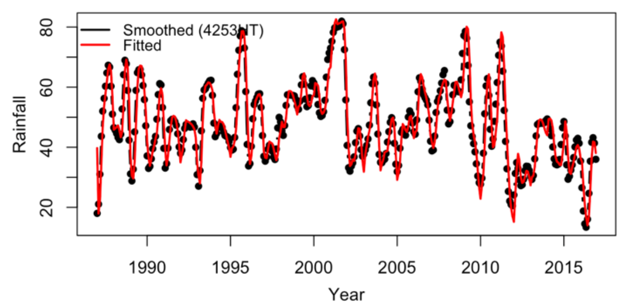


Figure 14 Smoothed data (4253HT) and fitted line (Holt-Winter method)

Table 4 shows the measures of central tendency of monthly maximum rainfall received from Jan 1987 to Dec 2016. Throughout this period, the highest rainfall distribution was recorded in Feb 2009, which is 176 mm, with a mean of 50.72 mm. The value 0.00 in the table indicates no rainfall received. Moreover, there were times where the maximum rainfall distribution received more than 100 mm, which are in: Oct 1987, Feb and Nov 1988, Nov 1992, Apr 1996, Dec 2000, Apr 2001, Jun 2003, Sep 2005, Nov 2007, Jul and Dec 2008, Jun and Nov 2010, Mar, April and Aug 2011, Dec 2014.

Table 4 Descriptive Statistics for monthly maximum rainfall (in mm) from January 1987 – December 2016

Mean	Max	Min	Median
50.72	176.00	0.00	47

The amount of rainfall in Malaysia is classified into four levels according to the Department of Drainage and Irrigation, Malaysia, which can be referred in Table 5. According to the table, light rainfall is from 1 to 10 mm, moderate rainfall is between 11 and 30 mm, heavy rainfall falls from 31 to 60 mm, while it is considered as very heavy rainfall if the amount exceeds 60 mm. Hence, these classes of rainfall level will be used as a reference in this study to discuss further on forecasted rainfall values.

Table 5 Classes of rainfall level in Malaysia

Class	Rainfall amount (mm)
Light	1-10
Moderate	11-30
Heavy	31-60
Very Heavy	>60

Table 6 shows the forecasted and actual monthly maximum rainfall in year 2017. Based on Table 5, all forecasted maximum rainfall can be categorized as heavy rainfall for the whole year. Besides, the highest amount of heavy rainfall in 2017 is predicted to be 41.722 mm in September. Moreover, there are accurate predictions found in February and June, as the predicted values of maximum rainfall received in both months are near to actual data. These results help people to plan properly any event to be held by considering the period of rainfall occurrences that have been predicted using this approach. However, the actual data states that very heavy rainfalls were received in January, April, July and August with 114, 68.5, 61.5 and 82 mm respectively. This contrasting result occurs due to uncontrollable factors that influence rainfalls distribution, such as climate changes. Besides, it is also due to the monthly maximum rainfall data used for this analysis which caused a huge variation in the data.

Based on the data provided by Department of Irrigation and Drainage Malaysia, the average of rainfall in Pahang is as tabulated in Table 7. It shows that the rainfall average in 2017 is quite higher than the previous years. This is influenced by several phenomena such as El Nino and La Nina which had affected the annual rainy seasons globally and caused the actual rainfall data in 2017 to increase slightly from the predicted rainfall data.

Table 6 Result from forecasting and the actual data of monthly maximum rainfall in year 2017

Month	Predicted	Actual
January	31.0140	114.0
February	**30.7974**	**33.0**
March	31.9907	44.0
April	34.8071	68.5
May	37.4848	51.5
June	**38.5151**	**38.5**
July	39.5999	61.5
August	40.7400	82.0
September	41.7220	37.5
October	41.6721	49.0
November	39.9365	34.5
December	36.4547	41.5

Table 7 Means of rainfall distribution in Pahang from 2013 until 2017

Year	2013	2014	2015	2016	2017
Mean	2327	1981	1591	1589	2633

Overall, the forecasted values of rainfall level obtained are near to the actual values of rainfall level, which is heavy and very heavy rainfall level respectively. These results help people to plan properly any event to be held by considering the period of rainfall occurrences that have been predicted using this approach.

4.0 CONCLUSION

This study used 4253HT smoother as one of the smoothing techniques to test on the sinusoidal signal with five levels of contamination via simulation. This study also compares the performance of 4253HT smoother with three different types of hanning which are Tukey [7], Shitan and Vazifean [15], and Husain [16]. The results obtained proved that the combination of 4253HT smoother and Shitan and Vazifean's hanning is capable to perform well to overcome three levels of contamination which are 10%, 25% and 75%. Besides, the performance of 4253HT was assessed using a real data related to rainfall, where the data only included the maximum values of rainfall distribution recorded monthly for a period of 30 years. At the first stage, the smoothed data gives a lower value of RMSE than the raw data. This indicates that the trend line formed using smoothed data is more appropriate and reliable for forecasting.

The performance of 4253HT was further studied in forecasting using smoothed values and raw data. As a result, the values of sigma, AIC and MAPE using smoothed values in forecasting are lower than the values obtained using raw data.

Based on the evidence attained from this study, the performance of 4253HT smoother remains excellent even with extreme value data. For future work, it is recommended to (1) use other types of smoothers such as 3RSSH to compare their performances, (2) apply ARIMA which is an advanced method in forecasting, integrated with 4253HT, (3) use the latest data that up to year 2021 for a better forecasting, and (4) use daily data instead of monthly maximum data for a better data representation and more reliable forecasting result.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Acknowledgement

We wish to thank Department of Irrigation and Drainage for providing the data (rainfall data) for this study, as well as the management teams and staff for their cooperation to allow this study to be conducted.

References

[1] Pitas, I. and A. N. Venetsanopoulos. 2013. *Nonlinear Digital Filters: Principles and Applications*. Springer Science & Business Media.

[2] Pitas, I. and A. N. Venetsanopoulos. 1990. *Nonlinear Digital Filters: Principles and Applications*. Dordrecht, The Netherlands. Kluwer.

[3] Bovik, A., T. Huang and D. Munson. 1983. A Generalization of Median Filtering using Linear Combinations of Order Statistics. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 31(6): 1342-50.

[4] Hird, J. N. and G. J. McDermid. 2009. Noise Reduction of NDVI Time Series: An Empirical Comparison of Selected Techniques. *Remote Sensing of Environment*. 113(1): 248-58. Doi: https://doi.org/10.1016/j.rse.2008.09.003.

[5] Conradie, W., T. De Wet and M. D. Jankowitz. 2009. Performance of Nonlinear Smoothers in Signal Recovery. *Applied Stochastic Models in Business and Industry*. 25(4): 425-44. Doi: https://doi.org/10.1002/asmb.774.

[6] Sargent, J. and A. Bedford. 2010. Improving Australian Football League Player Performance Forecasts using Optimized Nonlinear Smoothing. *International Journal of Forecasting*. 26(3): 489-97. Doi: https://doi.org/10.1016/j.ijforecast.2009.10.003.

[7] Tukey, J. 1977. *Exploratory Data Analysis*. Addison-Wesley Publishing Company Reading, Mass. Doi: https://doi.org/10.1002/bimj.4710230408.

[8] Velleman, P. F. and D. C. Hoaglin. 1981. *Applications, Basics, and Computing of Exploratory Data Analysis*. Duxbury Press.

[9] Rabiner, L., M. Sambur and C. Schmidt. 1975. Applications of a Nonlinear Smoothing Algorithm to Speech Processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 23(6): 552-7. Doi: https://doi.org/10.1109/TASSP.1975.1162749.

[10] Martin, R. and S. Schwartz. 1971. Robust Detection of a Known Signal in Nearly Gaussian Noise. *IEEE Transactions on Information Theory*. 17(1): 50-6.

[11] Masreliez, C. and R. Martin. 1977. Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter. *IEEE Transactions on Automatic Control*. 22(3): 361-71. Doi: https://doi.org/10.1109/TAC.1977.1101538.

[12] Azmi, N. N. K., M. B. Adam and N. Ali. 2018. Modified Compound Smoother in Median Algorithm of Span Size 12. 2: 4253HT.

[13] Velleman, P. F. 1975. Robust Nonlinear Data Smoothers: Theory, Definitions, and Applications. Ph. D. Thesis.

[14] Jin, Z. and B. Xu. 2013. A Novel Compound Smoother—RMMEH to Reconstruct MODIS NDVI Time Series. *IEEE Geoscience and Remote Sensing Letters.* 10(4): 942-6. Doi: https://doi.org/10.1109/LGRS.2013.2253760.

[15] Shitan, M. and T. Vazifean. 2011. *Exploratory Data Analysis for Almost Anyone*. UPM Press.

[16] Husain, Q. N. 2017. *Modifications of Tukey's Smoothing Techniques for Extreme Finance Data*. Universiti Putra Malaysia.

[17] Muntashir-Al-Arefin and P. M. A. Al. 2015. Package 'sleekts'. 4253H, Twice Smoothing. Compute Time series Resistant Smooth 4253H, Twice Smoothing Method.

[18] Velleman, P. F. 1980. Definition and Comparison of Robust Nonlinear Data Smoothing Algorithms. *Journal of the American Statistical Association*. 75(371): 609-15. Doi: https://doi.org/10.2307/2287657

[19] Jankowitz, M. D. 2007. *Some Statistical Aspects of LULU Smoother*. South Africa. Stellenbosch University.

[20] Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer. Doi: http://dx.doi.org/10.1007/978-1-4471-3675-0.

[21] Koehler, A. B., Snyder, R. D., and Ord, J. K. 2001. Forecasting Models and Prediction Intervals for the Multiplicative Holt-Winters Method. *International Journal Forecast*. 17(2): 269-286. Doi: https://doi.org/10.1016/S0169-2071(01)00081-4.

Appendix

Main R codes for 4253HT

```
z4 = NULL
z4[1] = y[1]
z4[2] = y[2]
z4[N] = y[N]
for (i in 3:(N-1)) {
  z4[i] = median(c(y[i-2], y[i-1], y[i], y[i+1]))}

z2 = NULL
z2[1] = y[1]
z2[2] = y[2]
z2[N] = y[N]
z2[N-1] = y[N-1]
for (i in 3:(N-2)) {
  z2[i] = median(c(z4[i], z4[i+1]))}

z5 = NULL
z5[1] = y[1]
z5[2] = y[2]
z5[N] = y[N]
z5[N-1] = y[N-1]
for (i in 3:(N-2)) {
  z5[i] = median(c(z2[i-2],z2[i-1],z2[i],z2[i+1],z2[i+2]))}

z3 = NULL
z3[1] = y[1]
z3[N] = y[N]
for (i in 2:(N-1)) {
  z3[i] = median(c(z5[i-1],z5[i],z5[i+1]))}
```