American Journal of Applied Mathematics and Statistics, 2022, Vol. 10, No. 3, 95-102
Available online at http://pubs.sciepub.com/ajams/10/3/4
Published by Science and Education Publishing
DOI:10.12691/ajams-10-3-4

A Comparative Study on Methods to Handle Wavelet Edge Distortion

AWSY Wickramasinghe, WMND Basnayake*

Department of Statistics, University of Colombo, Sri Lanka
*Corresponding author: nadeeka@stat.cmb.ac.lk

Received October 10, 2022; Revised November 17, 2022; Accepted November 27, 2022

Abstract The discrete wavelet transformation (DWT) is of considerable use in the domain of time series analysis. A fundamental problem in DWT is the distortions occurring at the edges while utilizing finite-length series. Insufficient information at boundary regions can lead to questionable accuracies of the transformation at the edges which will thus make a profound effect on further applications. Since the generally used methods to handle edge distortion such as zero padding has their own drawbacks, there are studies that have been done to alleviate the problem using mathematical techniques. With the main objective of finding an evidence-based strategy to reduce the edge effect inherent to DWT using statistical terminologies, this research compares the effect of statistical and machine learning-based denoising and extrapolating techniques in reducing-edge distortion using daily catchment flow series. The most suitable mother wavelet function and the decomposition level for the given series were considered as "biorthogonal 3.1" and level 2 and the edge effect was quantified using MAPE metric. The extrapolating techniques outperformed the denoising methods resulting Vanilla LSTM model with the lowest MAPE values and according to the averaged results taken considering 10 different points of the series, the Vanilla-LSTM and the SARIMA-LSTM hybrid model convincingly alleviate the edge distortion of all coefficients in a more generalized manner.

Keywords: *discrete wavelet transformation, neural networks, LSTM, time series*

Cite This Article: AWSY Wickramasinghe, and WMND Basnayake, "A Comparative Study on Methods to Handle Wavelet Edge Distortion." *American Journal of Applied Mathematics and Statistics*, vol. 10, no. 3 (2022): 95-102. doi: 10.12691/ajams-10-3-4.

1. Introduction

One frequently used way of collecting experimental data is as a sequence of values at successive equally spaced intervals in time. These sequences are called time series and the fundamental problem with time-series data is processing them to extract accurate and meaningful information. Practical time series are mostly complex in nature, noisy and strongly non - stationary. According to [1], non - stationary time series cannot be analyzed by using correlation methods or frequency-domain representations such as classical Fourier transformation. Usually, the most distinguishable information of a time series is captured by the frequency component of a series. The information which cannot be observed in the time domain can be observed in the frequency domain. But on some occasions, having both time and frequency domain information at once can become necessary and meaningful. A concept in the time-frequency domain that takes information in time and frequency simultaneously into account will be necessary for such instances.

To overcome this problem in analyzing non - stationary time series, Gabor [2] introduced the short-time Fourier transform. The idea behind it was to break a series into sub-samples and then apply Fourier transformation into each sub-sample. The one major drawback of this method was, using the same window to analyze the entire series. Hence, the frequency resolution becomes similar across the entire series. As an alternative to the above approaches, wavelet analysis has been proposed.

Wavelet analysis was originally invented to in seismic signal analysis and it has a higher resolution in both frequency and time domain [3]. The idea behind it first came to the spotlight in the late 1970s when J. Morlet, a geophysical engineer, came up with an idea to analyze different frequency bands with different window functions. These windows had compact support in both time and frequency domains. The honor of developing a wavelet with better localization properties goes to Alfred Haar, a German mathematician who discovered the first orthonormal wavelet basis function in 1909 [4].

The wavelet transform is the decomposition of a series into a set of basis functions consisting of expansions, shrinking and translations of a mother function called a wavelet [3]. Wavelet transformation is about looking at different scales and resolutions of a series. It can view a series in multiresolution and multi-scaled view and that property makes it outstanding among other transformations.

There are two types of wavelet transformations known as discrete wavelet transformation and continuous wavelet transformation. Discrete wavelet transformation has become a standard tool for time series and signal processing applications of several areas in research and industry. It can be described as an implementation of the wavelet transformation by using a set of discretized and restricted wavelet scales and translations. Wavelets for the transformation are generated by a single basic wavelet called the mother wavelet function by scaling and translation [5].

The procedure in wavelet analysis is to adopt the wavelet prototype function, mother wavelet; thus, the original series can be represented in terms of wavelet coefficients [6]. A wavelet can be described as a mini-wave-like oscillation with an amplitude that begins at zero. If one can select the most suitable mother wavelet function adapted to the data series, data can be sparsely represented and this makes wavelets an outstanding tool in various applications such as data compression [7].

A wavelet in wavelet transformation can be described as a mini-wave that is localized in time with a finite length. Each type of wavelet has a different shape, smoothness, and compactness and is useful for different purposes. Wavelets are limited in time and frequency. Hence, instead of lasting forever like a wave having no limit in time, it dies out quickly. This time-limiting quality provides wavelets with more resolution in the time domain.

The time localized property of wavelets allows wavelet transformation to obtain time/location information of a series. Hence, a signal can be multiplied with the wavelet at different locations in time. This procedure starts at the beginning of the signal and slowly moves towards the end of the signal. This is known as convolution. In most real-world applications, a particular series may have a finite support. As the wavelets get closer towards the end of the signal, computing convolution becomes harder since it has to consider non – existent values beyond the boundary [8,9,10].

Insufficient information in the boundary regions can lead to questionable accuracies of wavelet transformation at the edges of a series. The impacts of boundary effects influence extensively for some systems which have longer period sequences and thus require a high resolution in the frequency domain [11]. This is called edge distortion and can affect the accuracy of the transformation at the edges of the series.

If explained more comprehensively, the discrete wavelet transformation of the edge of a finite length series can behave in a different and complicated way rather than the usual transformation. This can be witnessed by splitting a series from the middle of the series and applying the discrete wavelet transformation. The transformation at the split point differs from the transformation of the original series at the same point. This happens because the split series do not have supporting observations for the transformation after the edge. But the original series will not have that problem at the said point since it's continuous.

This study proposes a novel methodology to handle the adverse effects that causes due to wavelet edge distortion through a comprehensive scientific comparison among potential methods.

2. Literature Review

According to Su, Liu, & Li, [12], "Boundary effects are caused by incomplete data in the boundary regions when the analysis window gets closer to the edge of a signal." Strang & Nguyen [13] has explained that border distortion is a result of choosing the wrong extension method when computing the convolution of finite length signals (p.339).

Edge effects have been defined using different names such as edge distortion and boundary effect by researchers in their studies [12,14]. According to Mallat [15], the edges of a signal or a series have to be treated differently than the other parts of the signal to avoid distortions. There are two alternative approaches to handle border effects such as, accepting the loss and truncating those questionable results or artificial extensions at the end of the boundaries [12].

There are various basic methods that are proposed to reduce the edge distortion such as zero padding, circular convolution, and symmetric extension [13]. These common extension methods can have their own drawbacks [12] and circular convolution and symmetric extension methods can lead to undesirable edge effects [16]. Basnayake et al., [14] have done a study on an algorithm that can handle the non-stationarity and non-linearity of a time series by building expert models to different data segments to avoid the forecasting error due to non-stationarity.

This study has proposed wavelet-based denoising to address the edge distortion that arises when using the discrete wavelet transform technique to decompose the real-time series into different data segments. The "Heuristic Sure" threshold that uses a level-dependent estimation of level noise has been selected as the most suitable wavelet threshold to denoise the time series. The wavelet denoising based algorithm has been proven effective in forecasting nonlinear and non-stationary time series data.

Williams & Amaratunga [16], have developed a discrete wavelet transformation technique that does not show the evidence of edge distortion by extrapolating a series. This study suggests a mathematical extrapolation method by determining the coefficients of a best-fit polynomial for the data points which are considered. The proposed method is proven to perform better than the circular convolution-based discrete wavelet transform and can be applied to orthogonal wave bases.

A study was done by Su et al., [10] that discusses the problem of having border effects in time-frequency analysis when using wavelet transformation. This study proposes a smooth extension to the series using Fourier transformation to avoid distortion. This method can preserve the waving characteristics of the time series which makes it more ideal than an alternative technique. According to the final results of the study, this extension scheme provides effective performance and the author suggests using this method in other time-frequency analysis techniques other than the wavelet transformation which exhibits border effects.

The discrete wavelet transformation is in considerable practical usage for applications related to sequential data. Utilization of such applications relies on the accuracy and the precision of the transformation and it could become problematic if the transformation results in questionable accuracies. Traditional methods which can be used to handle edge distortion are highly controversial and come with its own drawbacks. Thus, mathematicians and researchers have done several past studies to handle the edge effect and many of these approaches are mathematically supported. Mathematical approaches being deterministic in nature can always produce the same output from a starting condition or an initial state. Hence, a statistical approach that considers the native randomness in non – stationary time series would be beneficial and more effective for handling edge distortion in discrete wavelet transformation. Hence, this research is focused on using denoising and extrapolating techniques to handle edge distortion in DWT of time-series data, using both statistical and machine learning terminologies.

3. Methodology

This study uses the daily catchment flow data of Mahaweli reservoir from 1996 to 2015 which was obtained from the Mahaweli Authority. Mahaweli hydropower complex associated with the Mahaweli reservoir has six major power stations with a capacity of 666MW. The power stations under the Mahaweli hydropower complex are Victoria, Kotmale, Randenigala, Ukuwela, Rantambe, Bowatenna and Nilambe. Original data for the catchment flow to the upmost reservoir of Kothmale have been collected daily. These daily time series consist of 7305 observations with no missing values.

The first orthonormal wavelet basis function is the Haar wavelet which was introduced by the German mathematician Alfred Haar in 1909 [4].

The wavelet function can be defined as follows by ψ.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}.\psi\left(\frac{t-b}{a}\right) \tag{1}$$

Here a is the scale and b is the translation where t is time. Scaling is for either dilating or compressing the series. Low frequencies are represented by dilated or stretched scales and high frequencies are represented by small or compressed scales. The translation is related to the location of the mother wavelet and shifting it in time [17].

Discrete wavelet transformation is conducted by filtering the signal at different cut-off frequencies and different scales. It analyses a series by decomposing it into separate detailed and approximation coefficients. Both high and low pass filters are used in this process to analyze high and low frequencies respectively [17].

There are several ways to define discrete wavelet transformation mathematically. The DWT of a continuous series x(t) can be defined as,

$$DWT(m,p) = \int_{-a}^{+a} x(t).\psi_{m,p}\, dt \tag{2}$$

Where ψ(m,p) is the wavelet function which is created from the scales and translated mother wavelet function using scale m and translation p. Hence ψ(m,p) can be written as follows [18].

$$\psi_{m,p}(t) = \frac{1}{\sqrt{a_0^m}}.\psi\left(\frac{t - pb_0 a_0^m}{a_0^m}\right) \tag{3}$$

The intuition behind DWT is explained by Polikar, [17] as follows:

Consider a series x[n]. The procedure starts with sending the series x[n] through a half-band digital low pass filter with the impulse response g[n]. Filtering a series according to the mathematical operation known as the convolution of the series with the impulse response and can be defined as follows.

$$x[n]*g[n] = \sum_{k=-\infty}^{\infty} x[k]\cdot g[n-k] \tag{4}$$

Half band low pass filters remove all the frequencies higher than half of the highest frequency of the series. Since half of the frequencies are removed after sending a series through a low pass filter, half of the samples can be eliminated according to Nyquist's rule. Discarding samples of the series will sub-sample the series by two hence the scale is doubled now. Low pass filtering removes information that has high frequencies keeping the scale unchanged. But the sub-sampling process changes the scale of the series.

The signal is decomposed simultaneously by using a high pass filter h[n]. The filtering and sub-sampling processes of two filters add up to one level of decomposition. The low pass filters output the approximation coefficient while the high pass filters output the detail coefficient. This filtering procedure continues with the number of decomposition levels one would state. A level two decomposition can be represented as in Figure 1,

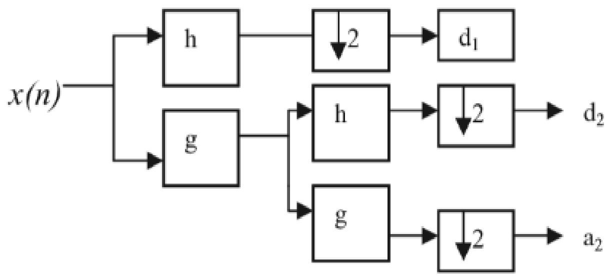

Figure 1. Level 2 decomposition in DWT

Here h and g are high and low pass filters respectively and d1 and d2 are the first and second level detail coefficients whereas a2 is the second level approximation coefficient. This process can be continued until a desired number of levels have been reached. The original series can be reconstructed using approximation and detailed coefficients by upsampling by two and passing through high and low pass synthesis filters and adding them [19].

Discrete wavelet transformation produces several detail and approximation coefficients depending on the decomposition level provided. The edge effect can be identified in each of these components separately. The difference between the transformations of the original series and the distorted series can be captured as the edge effect.

Most researchers prefer unit free measures for comparing time series. One such method is the mean absolute percentage error (MAPE) which calculates the error as a percentage of the actual value and is widely used to compare the error of the forecasts [20]. MAPE can be calculated as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{5}$$

where A_t is the actual value and the F_t is the forecasted value, n is the number of time points and t represents the time. This can be adapted to calculate the edge effect by taking the original series as the actual series and distorted series as the other series. Then the error percentage can be measured among the two series considering the same time step.

Single exponential smoothing is the simplest form of exponential smoothing. This technique can only be used for non-seasonal time series which do not show any specific trend. Many real-world time series consist of trends and seasonality, but these can be measured and removed to produce more stationary series [21]. Single exponential smoothing consist of one single parameter called smoothing parameter α. Considering a specific value of α, one–step forecasts can be iteratively produced through the series and then the mean square error can be calculated for each considered value of α. The α value which produces the least MSE value can be considered as the optimal smoothing parameter for a given time series [22].

Holt-Winters method is another smoothing technique which is also known as the triple exponential smoothing. It has the ability to handle data with seasonal variations and trend components [21]. In the triple exponential smoothing model, there are three parameters α, β and γ which are smoothing factors for level trend and seasonality. Other than that, there are other parameters that explain the type of time-series properties. They are, trend type, seasonality type, dampen type, damping coefficient and seasonal period.

Time series extrapolating also known as time series forecasting is used in many domains to gain future information in a useful manner. Chris Chatfield, [23] explains that "Univariate forecasts are methods where predictions depend only on the present and past values of the single series being forecasted, possibly augmented by a function of time such as a linear trend."

Time series forecast models need to learn past and present patterns of data to accurately predict values over time. Deep learning models have the ability to learn from raw and imperfect data and extract important features to improve forecasting [24]. LSTMs are a special type of RNN deep learning model that is designed to handle the long-term dependency problem. They have the ability to handle it with the vanishing gradient problem [25]. LSTMs have gates that regulate the flow of information through the network. These gates decide which information should be passed and which are not to the cell state which is the memory of the network.

Auto-Regressive Moving Average (ARMA) models are important in modeling stationary series. Auto-Regressive Integrated Moving Average (ARIMA) which is an extended version of ARMA models can process non-stationary series by differencing finitely many times. Trend patterns can be removed by differencing a time series hence ARIMA(p,d,q) models can be used to represent non-seasonal data with a trend [26]. If the time series consists of seasonal patterns, seasonality also should be modeled using an additional term. This is a special case of ARMA models known as Seasonal ARIMA (SARIMA).

SARIMA models can be represented using the notation as SARIMA(p,d,q)×(P,D,Q)s. Here, p is the trend AR order, q is the trend MA order and d is the trend difference order. There are four additional terms to model the seasonality where P is the seasonal AR term, Q is the seasonal MA term, D is the seasonal difference order and s is the seasonal period.

A hybrid approach using ARIMA and ANN models has been proposed by [27], for the purpose of time series forecasting. According to the results in this study, a hybrid approach using ARIMA and ANN models has been proven effective in forecasting time series rather than the existing models.

A complex time series can be further decomposed as linear and non-linear components using decomposing techniques. The linear component can be modeled by econometric models such as SARIMA for seasonal series. ANNs can model complex non-linear relationships hence can be used to model the non-linear component. Among ANN models, RNNs and LSTMs have the ability to model sequential data. The LSTM model can handle long-term dependencies rather than the RNN models. Hence, LSTM models will be used to model the non-linear component.

A time series can be either additive or multiplicative.

Additive model:

$$y_t = L_t + N_t \tag{6}$$

Multiplicative model:

$$y_t = L_t \times N_t \tag{7}$$

Suppose Lt represents the linear component and N_t represents the non-linear component. The model building procedure is as follows.

Consider a time series y_t, t = 1,2,3,…. A series of forecasts $\hat{L}_t$ can be generated by modeling the linear component using ARIMA/SARIMA models. Then a series of non-linear components (e_t) can be generated by comparing the actual value y_t with the forecasted value L $\hat{L}_t$.

In accordance with the additive model, $e_t = y_t - L_t$ and according to the multiplicative model, $e_t = y_t / L_t$. These are non-linear components and can be modeled using LSTM networks. Suppose the forecasts of the LSTM model is denoted by $\hat{L}_t$. Then the final results can be obtained by concatenating the two results. For an additive model, the final forecast will be,

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \tag{8}$$

For a multiplicative model, the forecast will be,

$$\hat{y}_t = \hat{L}_t \times \hat{N}_t \tag{9}$$

4. Results and Discussion

This study was conducted using a univariate time series which is daily catchment flow data from 1st January 1996 to 31st December 2015 that have been obtained from Mahaweli authority in Sri Lanka. The dataset consists of 7305 observations with no missing values.

According to the preliminary analysis, the catchment flow series is a series without a trend which consist of annual and biannual seasonal components. Considering the most dominant periodicity, priority was given to the annual seasonality. Hence, the annual seasonality will be considered in further analysis. The series was considered non-stationary without conducting formal tests, due to the presence of multiple seasonality.

As implied in literature, the edge distortion occurs as a result of having no information after the edge of a finite length series. Edge distortion in DWT of the catchment flow series can be observed as follows.

First, the original series was split at the 6012th observation considering it as the last observation of the series. From here onwards, the split catchment flow series means the series with 6012 data points whereas the 6012th is the last observation. The original catchment flow series represents the catchment flow series with 7305 observations.

Then the discrete wavelet transformation was applied to both original and the split catchment flow series separately. In the next step, each wavelet coefficient of the original and the split series were compared. The following figures denote those discrete wavelet coefficients of the aforesaid two series. In each figure, the coefficients of the original series are represented by the orange lines and the coefficients of the split series are represented by the blue lines.

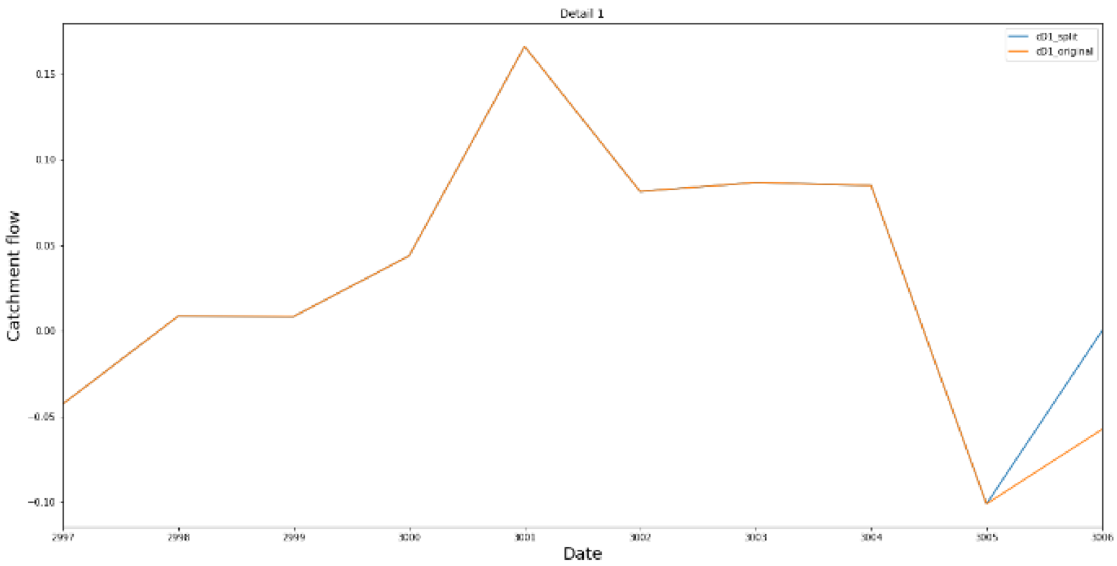

Figure 2. Detail 1 coefficients of original and split catchment flow series

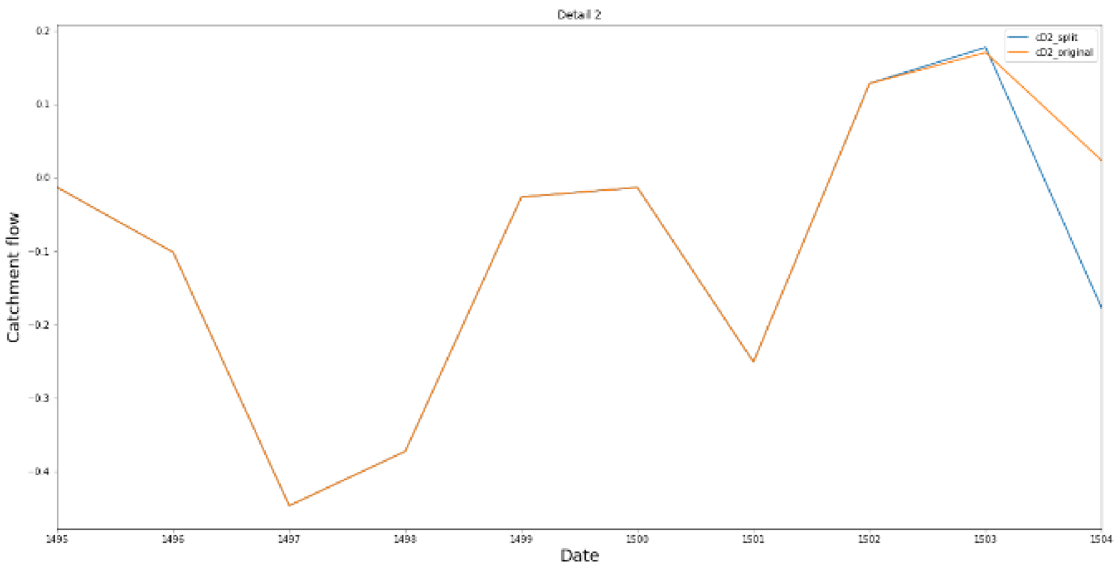

Figure 3. Detail 2 coefficients of original and split catchment flow series

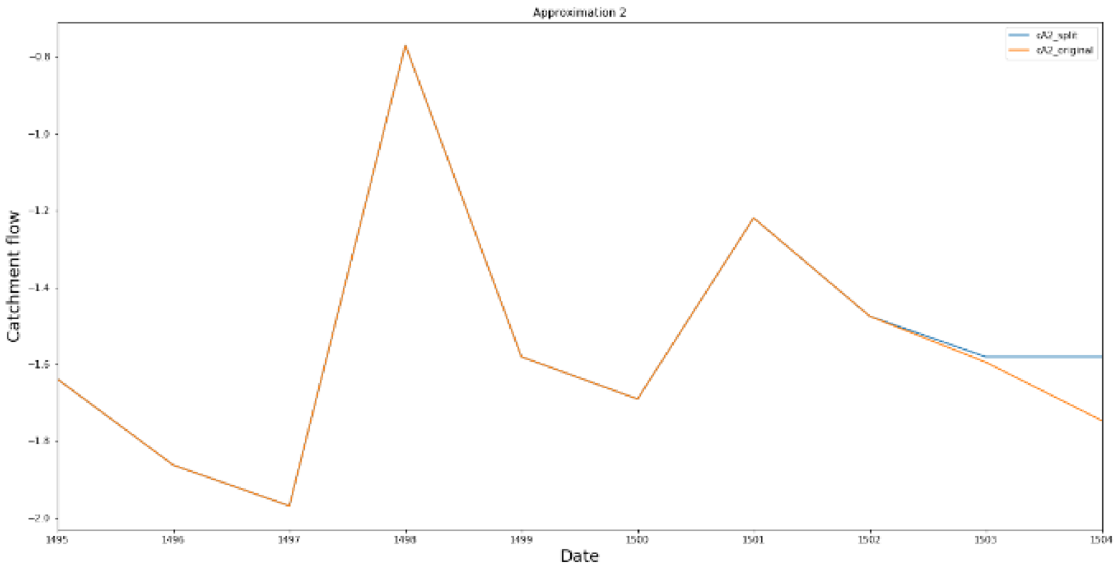

Figure 4. Approximation 2 coefficients of original and split catchment flow series

The split catchment flow series ends from the 6012th point and the original catchment flow series is longer than that with 7305 observations. If the discrete wavelet transformation at the edge of the split series is different from the transformation of the original series at the 6012th point, there is an edge distortion at the edge of the split catchment flow series due to the discontinuity in the series. The original catchment flow series transforms in the usual way at the 6012th point as it continues (does not end) at the considered point.

According to the above plots (Figure 2, Figure 3 & Figure 4), a distortion between the original and the split sequences can be observed in each coefficient. In each figure, the blue line deviates from the orange line which is the original catchment flow series. The distortion of the transformed split series from the transformation of the original series can be evaluated using the MAPE measure and the calculated MAPE values are as follows.

Table 1. MAPE of the edge effects

Discrete wavelet coefficient	MAPE (%)
Detail 1 coefficient	10.000
Detail 2 coefficient	85.655
Approximation 2 coefficient	1.035

The MAPE value captures the deviation of the blue line from the orange line. Detail 2 coefficient has the highest edge distortion whereas the approximation coefficient shows a slight deviation. Thus, it can be concluded that there is a distortion in the transformation of the split series from the transformation of the original series when the series is split at the 6012th observation. i.e. The transformation of the final series is distorted at the edge of the series.

(Note: Before conducting the discrete wavelet analysis, the most suitable mother wavelet function and the decomposition level must be decided. This has been calculated in a previous study done by [14] using the same catchment flow series used in this study. According to that, the most suitable mother wavelet function is “biorthogonal 3.1” and the decomposition level is 2. Hence, similar parameters were used in this study).

For this study, initially, seven methods were compared in reducing-edge effect considering 6012th observation as the last observation of the series under two major approaches which are denoising and extrapolating techniques. Finally, the same techniques were applied considering 10 other points as the last observation of the series and the edge effect was examined.

As the first approach, the denoising techniques were applied to handle the edge distortion. First, the original series (with 7305 observations) was denoised by using a statistical technique and transformed using the discrete wavelet transformation. Then the split series (with 6012 observations) was denoised by using the same statistical technique and applied discrete wavelet transformation. Then the distortion of the denoised transformed split series from the denoised transformed original series can be calculated using the MAPE metric. This is the calculated edge effect after denoising the series. This edge effect can be compared with the real edge effect in the catchment flow series at the considered point.

As the initial approach first the single exponential smoothing, and then the Holt-Winters method was applied to smooth the series. According to the results obtained by denoising the series using both methods, only the edge effect of detail 2 was reduced. As the denoising techniques did not come up with satisfying results, the need for an alternative approach was noticed.

As the next approach of the study, several statistical extrapolating approaches have been utilized in order to handle the edge distortion. In this approach, first, the split catchment flow series (with 6012 observations) was extrapolated using several statistical techniques. Then the discrete wavelet transformation was applied on both the original catchment flow series and the extrapolated catchment flow series. As the next step, the edge distortion was calculated.

According to the results of both extrapolation methods using SES smoothing and Holt-Winters smoothing, only the edge effect of detail 1 and 2 coefficients reduced As the above methods do not handle the edge effect as expected, the LSTM networks were used. The Vanilla LSTM and Bidirectional LSTM were used and among them, the Vanilla LSTM reduced the edge effect of all three discrete wavelet coefficients while the Bidirectional LSTM effectively reduced the edge effect of detail 1 and 2 coefficients only. Finally, a SARIMA-LSTM model was used to extrapolate the catchment flow series. This hybrid model reduced the edge effect in detail 1 and 2 coefficients only. A summary of the results obtained after applying the above-described methods is given in the table below.

Table 2. Summary of the MAPE of Edge Effects

	Method	MAPE (%)		
		D1	D2	A2
	Real Edge effect	10.00	85.65	1.03
Noise-reducing	Single exponential smoothing	19.85	69.78	1.69
	Holt winters	57.53	47.25	3.34
Forecasting	Single exponential smoothing	1.01	2.91	1.23
	Holt winters	0.25	18.69	1.83
	Vanilla LSTM	0.45	2.77	0.94
	Bidirectional LSTM	9.02	2.01	1.95
	SARIMA LSTM hybrid model	7.81	16.67	1.83

All the above methods were applied considering one location in the time series. However, the nature of the edge effect can vary according to the nature of the last observations of a series. Examining several instances can be less biased rather than sticking to only one instance. Hence, it was decided to apply the above-explained techniques contemplating the edge effects at different splits.

Suppose x is the last observation of the catchment flow series. The edge effect was calculated when x is equal to 6012, 6080, 6500, 6550, 6600, 6650, 6700, 6750, and 6850. Then, the above denoising and extrapolating techniques that were discussed in the advanced analysis were applied to examine the change in the edge effect of DWT. The change in edge effect at each instance was measured by using the MAPE criteria. Finally, the MAPE values of each statistical technique were averaged. The below table represents the averages of the MAPE values calculated considering each technique at different points of the series. According to the following average MAPE values, the noise reduction techniques are not as effective as extrapolating techniques.

Initially, the study was conducted considering the edge effect at one point of the series. Hence the following results will be discussed contemplating the results obtained using the finite split series where the 6012th observation was the last observation.

Outcomes of this study suggest that the single exponential smoothing performs far better as a denoising technique compared to the Holt-Winters method in the context of reducing the edge effect. Above denoising techniques reduced the edge effect of the detail 1 and 2 coefficients only. Moreover, the reduction of the edge effect is not as high as expected. These results do not explain a clear relationship between denoising and edge effect also.

Table 3. Summary of the average MAPE of Edge Effects

	Method	Average MAPE (%)		
		D1	D2	A2
	Real Edge effect	10	24.03	13.54
Noise-reducing	Single exponential smoothing	13.21	33.98	6.27
	Holt winters	31.70	52.53	8.86
Forecasting	Single exponential smoothing	22.49	15.81	16.28
	Holt winters	9.63	17.40	10.39
	Vanilla LSTM	9.21	14.16	6.57
	Bidirectional LSTM	23.69	19.24	11.88
	ARIMA – LSTM hybrid model	9.88	10.78	9.87

Denoising a series alters the resulting sequence. The discrete wavelet coefficients of a denoised series are far different from the discrete wavelet coefficients of the original series. However, the reduction of the edge effect can be measured as a percentage using MAPE and compared with the initial edge effect after denoising the series. On the other hand, denoising a series can produce meaningful information by removing unwanted fluctuations. Hence this technique has its own advantages and disadvantages. The effectiveness of using this technique to handle edge distortion depends on the purpose of the transformation.

When the series was extrapolated using SES and Holt Winter's method, the edge distortion in the resulting coefficients of the DWT seems to be reduced in detail coefficients 1 and 2. However, the forecasting error (MSE) of the results obtained by the SES method was 0.004 and the MSE of the Holt-Winters method was 0.006. Hence the model with the lowest MSE value performed better in handling the edge effect. Moreover, out of the two LSTM networks used, the Vanilla LSTM model performed considerably better than the bidirectional LSTM network. The edge effect of all three wavelet coefficients has been decreased after extrapolating the split series using Vanilla LSTM. The mean squared error of the forecasts given by Vanilla LSTM and bidirectional LSTM are 0.004 and 0.014 respectively. The model with the lowest forecasting error has shown the most effective results in reducing-edge effect. However, in this study, the forecasting error of the hybrid model is higher than the forecasting error given by the Vanilla LSTM model alone. Thus, the Vanilla LSTM has outperformed the hybrid model in reducing the edge distortion as well.

Almost all the extrapolation techniques have outperformed the noise reduction techniques. However, in general, the Vanilla LSTM excelled significantly in reducing-edge distortion and extrapolation using SES was also nearly more effective than the other techniques. Compared to the mean squared error of extrapolation, it is a fact that the reduction of the edge distortion of the wavelet transformation is directly proportional to the decrease in the MSE of the extrapolation. The edge effect can be reduced with the decrease in the extrapolation error.

According to the average MAPE values of the edge distortions recorded after applying the above techniques at several points of the series, the hybrid model and the Vanilla LSTM model have shown the most effective results in handling the edge distortion by reducing the edge effect in all three discrete wavelet coefficients. The Hybrid model had a moderate performance in the previous individual instance considered. However, according to the collective results, it has been able to reduce the edge distortion successfully resulting in lower average MAPE values. In most of the instances considered, extrapolating using SES is not as effective compared to the one split (at 6012th observation) considered earlier.

In both individual instances and collective instances, the Vanilla LSTM outperformed all the techniques.

5. Conclusion

Conclusions derived on completion of the research are listed as follows. Most effective models in reducing-edge distortion in discrete wavelet transformation have been found as vanilla LSTM and extrapolation using Single Exponential Smoothing. According to the averaged results, the Vanilla LSTM and Hybrid model can be used as extrapolation techniques to handle the edge distortion in a more generalized manner. The statistical denoising methods do not perform as expected in reducing the edge effect of discrete wavelet transformation. Both denoising techniques reduced the edge effect up to some extent and that also only in detail 1 and 2 coefficients. On average, the denoising techniques do not perform well in handling edge distortion.

References

[1] Boashash, B. (1987). *Theory, Implementation and Application of Time-Frequency Signal-Analysis using the Wigner-Ville Distribution*. Retrieved from https://www.researchgate.net/profile/Boualem_Boashash/publication/265752441_Theory_Implementation_and_Application_of_Time-Frequency_Signal_-Analysis_using_the_Wigner-Ville_Distribution/links/573f481508ae9f741b321c4b/Theory-Implementation-and-Application-o.

[2] Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, pp. 429-441.

[3] Daubechies, I. (1992). Ten Lectures on Wavelets. *Ten Lectures on Wavelets*

[4] Daubechies, I. (1996). Where do wavelets Come From? 84(4).

[5] Valens, C. (1999). A really friendly guide to wavelets. *Ed. Clemens Valens.*

[6] Kaiser, G. (2010). A Friendly Guide to Wavelets. In *Springer Science & Business Media*. Retrieved from https://books.google.lk/books.

[7] Varma, T., Chitre, V., & Patil, D. (2012). The haar wavelet and the biorthogonal wavelet trransforms of an image. *International Journal of Engineering Research and Applications*, 268-291.

[8] Cohen, A., Daubechies, I., & Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, *1*(1), 54-81.

[9] De Queiroz, R. L. (1992). Subband processing of finite length signals without border distortions. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *4*, 613-616.

[10] Su, H., Liu, Q., & Li, J. (2012). Boundary effects reduction in wavelet transform for time-frequency analysis. *WSEAS Transactions on Signal Processing*, *8*(4), 169-179.

[11] Addison, P. S. (2002). The illustrated Wavelet Transform Handbook. In *Biomedical Instrumentation and Technology*.

[12] Su, H., Liu, Q., & Li, J. (2013). Adaptive approach for boundary effects reduction in rotating machine signals analysis. *Advances in Mechanical Engineering*, *2013*.

[13] Strang, G., & Nguyen, T. (1996). Wavelets and Filter Banks. In *SIAM*. Retrieved from. (1996).

[14] Basnayake, W., Attygalle, M. D. T., Liyanage-hansen, L., & Nandalal, K. D. W. (2019). *Modified 1D Multilevel DWT Segmented ANN Algorithm to Reduce Edge Distortion*. *7*(1), 25-31.

[15] Mallat, S. (1999). Fourier kingdom. A Wavelet Tour of Signal Processing, 20-41.

[16] Williams, J. R., & Amaratunga, K. (1997). A Discrete Wavelet Transform Without Edge Effects Using Wavelet Extrapolation. *Journal of Fourier Analysis and Applications*, *3*(4), 1-19.

[17] Polikar, R. (2019). *the Wavelet Tutorial Second Edition Part I*. 1-67.

[18] Oliveira, M. O., & Bretas, A. S. (2009). Application of discrete wavelet transform for differential protection of power transformers. *2009 IEEE Bucharest PowerTech: Innovative Ideas Toward the Electrical Grid of the Future*.

[19] Chaovalit, P., Gangopadhyay, A., Karabatis, G., & Chen, Z. (2011). Discrete wavelet transform-based time series analysis and mining. *ACM Computing Surveys*, *43*(2).

[20] Armstrong, B. J. S., & Collopy, F. (1992). Error Measures For Generalizing About Forecasting Methods: Empirical Comparisons By J. Scott Armstrong and Fred Collopy Reprinted with permission form. *International Journal of Forecasting*, *8*(1), 69-80.

[21] Chatfield, C. (2003). The Analysis of Time Series: An Introduction. In *The Analysis of Time Series: An Introduction*.

[22] Kalekar, P. (2004). Time series forecasting using Holt-Winters exponential smoothing. *Kanwal Rekhi School of Information Technology*, (04329008), 1-13.

[23] Chatfield, Chris. (2000). *TIME-SERIES FORECASTING*.

[24] Längkvist, M., Karlsson, L., & Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, *42*(1), 11-24.

[25] Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: Continual prediction with LSTM. *IEE Conference Publication*, *2*(470), 850-855.

[26] Casella, G., Fienberg, S., & Olkin, I. (2006). Springer Texts in Statistics. In *Design* (Vol. 102).

[27] Wang, L., Zou, H., Su, J., Li, L., & Chaudhry, S. (2013). An ARIMA-ANN Hybrid Model for Time Series Forecasting Li. *Systems Research and Behavioral Science*, *30*(27), 244-259.

© The Author(s) 2022. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).