A Discrete Haar Wavelet Based Approach for Visualizing Error Regarding a Simulated Time Series

Ganegoda N.C., Kumara K.K.W.A.S. and Tantrigoda D.A., *University of Sri Jayewardenepura*
Boralugoda S.K., Perera S.S.N., *University of Colombo*

***Abstract*— Point-wise error between a time series and its simulated series is not a stronger technique in data analysis. However, extending this point-wise error into a decomposition of two different qualitative values is addressed here. The decomposition is facilitated by discrete Haar wavelet. Ultimately, a spectrum has been designed to illustrate the error allowing localized analysis and interpretations too.**

Keywords-decomposition; error; Haar wavelet; spectrum; time series

I. INTRODUCTION

Error calculation is an important aspect in numerical analysis. It ultimately benefits in applied mathematical contexts such as model validation too. Classical but still the most convenient way of error estimation is to determine just the difference between actual value and approximated value [1]. This is applicable in time series as well since one can execute point-wise error calculation. However, since time series data represent a longer trend, it is not reasonable to emphasis only on point-wise errors. In this communication, a spectrum has been proposed to illustrate the error associated with series of observed values (data) and their simulated approximations.

Discrete Haar wavelet is the underpinning technique for the proposed spectrum. Main conceptual rule in wavelet is to decompose a series of values into two sets namely scaling coefficients (father wavelets) and wavelet coefficients (mother wavelets). This decomposition allows recognizing pattern of the series in different scales of localization. Here, those decomposed values lead to make a spectrum to acquire the quality of the error rather than the quantity.

II. MATHEMATICAL FORMULATION

Here, Haar wavelet is used to decompose series values with $\frac{1}{2}$ times stretch for classical Haar matrix $\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$. First, data series $\{d_i\} = (d_1, d_2, d_3, \dots, d_{2^n})$ is transformed to a series of two-component vectors as

$$V = \big((d_1, d_2), (d_3, d_4), \dots, (d_{2^n-1}, d_{2^n})\big).$$

Here, n is a positive integer allowing number of time points is a power of 2. It facilitates to carry out required coupling in further levels. Thereafter, each vector in V is right multiplied by the matrix $H = \frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$ to obtain; $W = \big((a_1, b_1), (a_2, b_2), \dots, (a_{2^{n-1}}, b_{2^{n-1}})\big)$, where $a_i = \frac{d_{2i-1}+d_{2i}}{2}$ and $b_i = \frac{d_{2i-1}-d_{2i}}{2}$ for $i = 1,2,3,\dots,2^{(n-1)}$. Then series $\{a_i\}$ and $\{b_i\}$ are called scaling coefficients and wavelet coefficients respectively. In this application of wavelet, averaging is the scaling function and wavelet coefficient represents slope by taking length between two time points as 2 units. This process is continued for further levels by taking scaling coefficients $\{a_i\}$ as $\{d_i\}$ recursively. Then the second level takes average values in the first level as its data series and third level takes average values in the second level as its data series and so on. It can be smoothly carried out when the number of data is a power of 2 as taken above.

Next, series of simulated values $\{s_i\} = (s_1, s_2, s_3 \ \dots, s_{2^n})$ is also decomposed similarly as in W to have a series of two-component vectors $U = \big((p_1, q_1), (p_2, q_2), \dots, (p_{2^{n-1}}, q_{2^{n-1}})\big)$, where $\{p_i\}$ and $\{q_i\}$ represent scaling coefficients and wavelet coefficients. Now, the error decomposition is defined as follows.

Average error (scaling error) $\quad x_i = a_i - p_i$
Slope error (wavelet error) $\quad y_i = b_i - q_i$

Then error $e_{2i-1} = d_{2i-1} - s_{2i-1} = x_i + y_i$ and $e_{2i} = d_{2i} - s_{2i} = x_i - y_i$ for $i = 1, 2, 3, \dots, 2^{n-1}$. Thus, $\{x_i\}$ and $\{y_i\}$ form a decomposition on point-wise errors. Next, more responsible cause of error can be identified by determining the largest of $|x_i|$ and $|y_i|$ in each case as follows.

- If $|x_i| > |y_i|$, then average error is the dominant cause of error (denoted by A)
- If $|x_i| < |y_i|$, then slope error is the dominant cause of error (denoted by S)
- If $|x_i| = |y_i|$, then both average and slope error equally responsible (denoted by E)

Thereafter, the spectrum will be designed by A (dominant cause is average error), S (dominant cause is slope error) and E (both errors are equally responsible). After having error decomposition for all levels, a collective spectrum can be presented. Henceforth, it is named as Decomposed Error Spectrum using Wavelets (DES-W).

DOI: 10.5176/2251-3043_3.1.232

III. Results

To illustrate the context of DES-W, time series data for annual changes in global temperature (1970 – 1985) available in a database (Time Series Data Library) is used here [2]. A simulated series with 50% variation around the data is obtained to determine errors. Since data set has 16 entries, there are 4 levels in the corresponding DES-W according to the coupling. One trialed case has been presented below with graphs to allow better understanding. Following result in Theorem 1 is also useful to have a comparative idea.

Theorem 1: Let s_i be the simulated value for data d_i and $e_i = x_i + y_i$ be the error decomposition, where x_i and y_i be the average error and slope error respectively defined by $x_i = \left(\frac{d_i+d_{i+1}}{2}\right) - \left(\frac{s_i+s_{i+1}}{2}\right)$ and $y_i = \left(\frac{d_i-d_{i+1}}{2}\right) - \left(\frac{s_i-s_{i+1}}{2}\right)$. Then, in a data series plot along with corresponding simulated values, the line joining d_i and d_{i+1} and the line joining s_i and s_{i+1}

i. do not intersect if and only if $|x_i| > |y_i|$ (ie. average error is the dominant cause);
ii. intersect at one of the time points or coincide if and only if $|x_i| = |y_i|$ (ie. both average and slope errors are equally responsible) and
iii. intersect at a point other than two time points if and only if $|x_i| < |y_i|$ (ie. slope error is the dominant cause).

Proof:
The result can be easily proved using the following implication.

$$|x_i| \gtreqless |y_i|$$
$$\Updownarrow$$
$$\left|\left(\frac{d_i + d_{i+1}}{2}\right) - \left(\frac{s_i + s_{i+1}}{2}\right)\right| \gtreqless \left|\left(\frac{d_i - d_{i+1}}{2}\right) - \left(\frac{s_i - s_{i+1}}{2}\right)\right|$$
$$\Updownarrow$$
$$\left|\left(\frac{d_i - s_i}{2}\right) + \left(\frac{d_{i+1} - s_{i+1}}{2}\right)\right| \gtreqless \left|\left(\frac{d_i - s_i}{2}\right) + \left(\frac{s_{i+1} - d_{i+1}}{2}\right)\right|$$

Now, $|x_i| = |y_i|$ if and only if $\left(\frac{d_i-s_i}{2}\right) + \left(\frac{d_{i+1}-s_{i+1}}{2}\right) = \pm\left(\left(\frac{d_i-s_i}{2}\right) + \left(\frac{s_{i+1}-d_{i+1}}{2}\right)\right)$. By considering (+), it can be shown that $d_{i+1} = s_{i+1}$. For $(-)$, $d_i = s_i$ is the case. Coincidence of $d_i = s_i$ and $d_{i+1} = s_{i+1}$ yields $x_i = y_i = 0$, which is also an obvious possibility of $|x_i| = |y_i|$. These arguments claim the result (ii).

Next, if $d_i > s_i$ then $|x_i| \gtrless |y_i|$ if and only if $d_{i+1} \gtrless s_{i+1}$ and if $d_i < s_i$ then $|x_i| \gtrless |y_i|$ if and only if $d_{i+1} \lessgtr s_{i+1}$ which yield (i) by the cases of $|x_i| > |y_i|$ and (iii) by the cases of $|x_i| < |y_i|$ that conclude the proof of Theorem 1.

Theorem 1 suggests that if there is errors in trend (ie. pattern of increasing and decreasing), then slope error is the dominant type. If there are vertical shifts rather than the changes in pattern, then average error takes the dominance.

DES-W of the trialed case is as follows. Fig 1 is the respective plot.

Level 1 : ASSASSSS
Level 2 : SSAA
Level 3 : SS
Level 4 : A

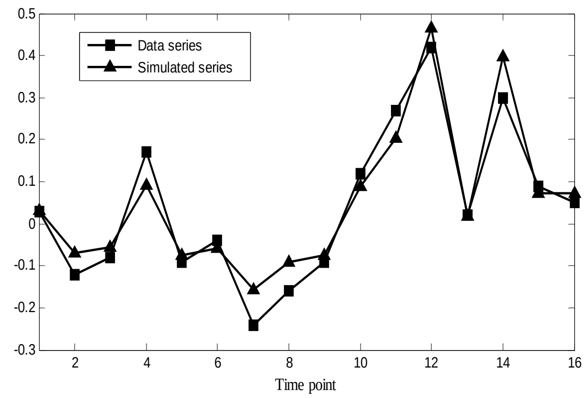

Figure 1: Data and simulated series of annual changes in global temperature (1970 – 1985).

Here, slope error is the dominant type in Level 1 indicating increasing and decreasing patterns are interchanged for data and simulation in respective coupling. In Level 2 where average values of Level 1 are accounted has slope errors first and then average errors. According to the localization there, first half is responsible with an error in trend and second half accounts more with a vertical shift in series values. This interpretation for Level 2 and for further levels is not directly visible in the graph which reveals a characteristic feature of DES-W. Level 3 spectrum suggests that average values of both first and last two quarters are more responsible with slope errors. Finally, Level 4 has the interpretation that average values in first and second halves are with a vertical shift rather than a difference in trend.

IV. Discussion

In data analysis, numerical measures such as mean, median, standard deviation and illustrative measures such as distribution plots, graphs show their appropriateness according to the user's requirements. However, spectrum approaches as introduced here have not yet been recognized up to that potential. Reading a spectrum is more like reading a graph where it allows seeing full outcome without restricting the analyzed situation to several numerical figures. Thus, spectrums are more on illustrative side bringing qualitative variation rather than quantitative outcome. On the other hand, comparing several spectrums may be more convenient than comparing graphs. Furthermore, in computational approaches, it is much easier to deal with a spectrum such as DES-W, which has illustrative layered details. Especially, this advantage along with the applicability of wavelet transforms could be seen in encoding and decoding processes regarding a large amount of information [3-4].

DES-W has the ability of providing localized information on error. Therefore, it is more suitable to use in phenomena with rapid or unexpected fluctuations such as weather patterns, disease transmission, financial transactions etc.. Applicability of wavelets emerges due to this strong localization ability. This feature is vastly discussed in literature for transforms from time domain to frequency domain [5]. Fourier transform is the main contrast, which retrieves the global frequency content. Hence it could be applied in designing tools to analyze smaller portions of a phenomenon via an illustration designed for full phenomenon. This potential is very much evident in wavelet applications used in image processing, where a vast amount of data must be manipulated [6-7].

At a glance, one may see a limitation due to the way of coupling series values. For instance, there are no comparison once second and third entries and fourth and fifth entries are coupled and so on. However, latter levels remove this drawback for certain extent as a result of scaling process. Further, such a pinpoint emphasis is not required once we deal with vast amount of data. If it is required to see the effect at first level, coupling process can be shifted by one entry provided that an additional entry is available.

In DES-W, the simplest wavelet type Haar is used to absorb the features of average and difference in the corresponding decomposition. One can modify it or use another wavelet type to design mathematical tools in data analysis with different bases of transformation. Those alterations would be formulated in a problem-specific manner and according to the end-users requirements.

REFERENCES

[1] S.D. Conte and C. De Boor, Elementary Numerical Analysis. 3rd ed. McGraw-Hill, 1965.

[2] R.J. Hyndman, Time Series Data Library [online] http://data.is/TSDLdemo. [Accessed: August 12, 2012].

[3] H.L. Cycon, M. Palkow, T.C. Schmidt and M. Wahlisch, A Fast wavelet-based video codec and its application in an IP version 6-ready serverless videoconferencing system [pdf], Available at < http://iphome.hhi.de/marpe/download/icwaa03.pdf> [Accessed: August 02, 2012].

[4] I. Maglogiannis, “Wavelet-Based Compression With ROI Coding Support for Mobile Access to DICOM Images Over Heterogeneous Radio Networks”, IEEE Transactions on Information Technology in Biomedicine, vol. 13:4, 2009, pp. 458 - 466.

[5] R.J.E. Merry, Wavelet Theory and Applications: A literature study[pdf], Available at <http://alexandria.tue.nl/repository/books/612762.pdf> [Accessed: August 20, 2012].

[6] K.H. Talukder and K. Harada, “Haar Wavelet Based Approach for Image Compression and Quality Assessment of Compressed Image”, IAENG International Journal of Applied Mathematics (Advance online publication), vol 36:1, 2007.

[7] S. Ye, Q. Sun and E. Chang, “Edge Directed Filter based Error Concealment for Wavelet-based Images”, IEEE International Conference on Image Processing, 2004, pp. 809-812.

Authors:

Ganegoda N.C., Kumara K.K.W.A.S.
Department of Mathematics
University of Sri Jayewardenepura
Nugegoda, Sri Lanka

Tantrigoda D.A.
Department of Physics
University of Sri Jayewardenepura
Nugegoda, Sri Lanka

Boralugoda S.K., Perera S.S.N.
Department of Mathematics
University of Colombo
Colombo, Sri Lanka