International Journal of Advances in Computer Science & Its Applications– IJCSIA
Volume 5: Issue 1 [ISSN : 2250-3765]
Publication Date : 30 April, 2015

Multiscale Analysis of Water Quality Time Series Data using the Hilbert Huang Transform

Adarsh S, Janga Reddy M

Abstract—**This paper presents the multiscale spectral analysis of four water quality time series data from an Indian river. First, the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) is employed for multiscale decomposition and the resulted orthogonal modes namely Intrinsic Mode Functions (IMFs) are subsequently subjected to the Normalized Hilbert Transform (NHT). The spectral representation clearly depicts the nonlinearity and non-stationarity of the datasets and the time varying behavior of dominant frequency. The marginal Hilbert spectrum of different parameters shows that the dominant frequency of most of the pollutants is at high frequency range which indicates the significant anthropogenic impacts in the study area. Also the trend analysis performed upon the instantaneous amplitudes show that the high frequency components are responsible for overall trend of the four time series during the study period under consideration. The multiscale decomposition process and the results of spectral analysis may improve the modeling efforts on the river system.**

Keywords*—*Amplitude, Frequency, Multiscale Decomposition, Hilbert Huang Transform, Water Quality

I. Introduction

Water quality modeling is of great concern among the environmental modelers and hydrologists for ensuring the quality standards of natural water bodies. Most of the water quality time series data possess nonlinear and non-stationary characteristics and a multiscale spectral analysis study may give more insight to the modeling practices. Hilbert Huang Transform (HHT) proposed by Huang et al. [1] is one of the spectral analysis tools which overcomes the limitations concerning the '*a priori*' selection of decomposition levels, basis functions, linearity and stationarity requirements of the time series, lack of ability for local scale characterization etc. possessed by traditional spectral analysis techniques such as Fourier Transform, Wigner Ville Distribution, Wavelet Transforms. The use of HHT has been gaining popularity in geophysical applications [2], hydrology [3-6] as most of the hydrological time series possess nonlinearity and non-stationarity characteristics.

For the analysis of water quality time series data, only a few applications of HHT has been reported sofar [7-10] and none in the Indian context. Also, performing a formal trend analysis of spectral amplitudes is an important step in the analysis of water quality time series. The traditional Empirical Mode Decomposition (EMD) may sometimes possess a limitation such as existence of modes in multiple frequency ranges (popularly known as 'mode mixing') which in turn can lead to misinterpretation of spectra. Thus to avoid such problems, noise assisted and ensemble averaged variants of EMD were proposed in recent past [11, 12]. The traditional Hilbert Transform may sometimes lead to instantaneous frequencies which are mathematically incorrect or with less physical meaning. To circumvent such limitations, Normalized Hilbert Transform (NHT) and direct quadrature schemes were proposed by Huang et al. [13]. Therefore the specific objectives of this paper are framed as follows : (i) to propose a novel method for the spectral analysis of water quality time series by conjunctive application of a variant of EMD namely the CEEMDAN and NHT (ii) to estimate the dominant frequencies by the construction of marginal Hilbert Spectrum (MHS) and (iii) to perform a trend analysis of spectral amplitudes of the series at different scales.

II. Hilbert Huang Transform

Hilbert Huang Transform is a popular method for analyzing nonlinear and non-stationary time series data. The HHT method involves two phases (i) a multiscale decomposition of time series to get orthogonal series called Intrinsic Mode functions (IMFs) (ii) the NHT of the obtained IMFs. A typical EMD process contains (i) the identification of local extrema points and fitting of spline functions connecting them (ii) compute the mean series (of extrema) and find a residue series (obtained by subtraction of mean from the original series). The above process is known as 'sifting' and the sifting operation is iteratively continued till the resulting series become a zero mean series with total number of extrema differ from the summation of number of local maxima and local minima points at the most by one. The process is repeated till the final residue series is monotonic or contains only one peak. A detailed description on the EMD can be found in Rao and Hsu [7]. The IMFs obtained sometime show 'mode mixing' and to overcome this issue, Torres et al. [12] proposed the CEEMDAN algorithm which is rarely applied for environmental time series. The important steps of the CEEMDAN algorithm are presented below :

ADARSH S
Research Scholar, IIT Bombay, Maharashtra
Assistant Professor, TKM College of Engineering, Kollam, Kerala
India

Dr. JANGA REDDY M
Assistant Professor, IIT Bombay, Maharashtra
India

1. Perform the EMD for M realizations $X_m(t) = X(t) + \varepsilon_0 w_m(t)$ and compute the first mode of CEEMDAN $\overline{\overline{IMF_1(t)}}$, by averaging the realizations

$$\overline{\overline{IMF_1(t)}} = \frac{1}{M}\sum_{m=1}^{M} IMF_m(t) \qquad (1)$$

where, $m=1,2,\ldots,M$ the index for realizations; ε_o is the noise parameter for the initial step.

2. At the first stage (k = 1) calculate the unique first residue as : $R_1(t) = X(t) - \overline{\overline{IMF_1(t)}}$ (2)

3. Decompose realizations $R_{1m}(t) = R_1(t) + \varepsilon_1 E_1(w_m(t))$, $m = 1,2,\ldots.,M$ until their first EMD mode gets evolved. Then compute the second mode $\overline{\overline{IMF_2(t)}}$

$$\overline{\overline{IMF_2(t)}} = \frac{1}{M}\sum_{m=1}^{M} E_1[R_1(t) + \varepsilon_1 E_1(w_m(t)] \qquad (3)$$

where, ε_1 is noise parameter for stage 1 (k=1); the operator $E_k(.)$ is an operator represents the evolution of the k^{th} mode by EMD.

4. Calculate the k^{th} residue as

$$R_k(t) = R_{k-1}(t) - \overline{\overline{IMF_k(t)}} \qquad (4)$$

for $k=2, 3,\ldots, K$

where, $\overline{\overline{IMF_k(t)}}$ is the IMFs obtained by CEEMDAN

5. Compute the first EMD mode of $R_k(t) + \varepsilon_k E_k(w_m(t))$, and define the $(k+1)^{th}$ mode by CEEMDAN as

$$\overline{\overline{IMF_{k+1}(t)}} = \frac{1}{M}\sum_{m=1}^{M} E_1[R_k(t) + \varepsilon_k E_k(w_m(t)] \qquad (5)$$

6. Go to step (4) for next k

Steps 4 to 6 are performed until the obtained residue is no longer feasible to be decomposed (i.e., the residue is monotonic or having only one extrema).

$$R_K(t) = X(t) - \sum_{k=1}^{K} \overline{\overline{IMF_k(t)}}. \qquad (6)$$

The obtained IMFs are the most appropriate inputs to perform the Hilbert transform which may help to study the spectral characteristics of a nonlinear and non-stationary time series in the time-frequency domain. The mathematical background of HT can be found in Huang et al. [1] or Rao and Hsu [7]. But the traditional HT may sometimes lead to instantaneous frequencies that are of less physical meaning (such as negative frequency) or mathematically incorrect. To overcome such problems, Huang et al. [13] proposed a normalization scheme for the HT. The theoretical background of this scheme is available in [2, 13]. In this study the CEEMDAN based multiscale decomposition and the normalization scheme are conjunctively used for analyzing four water quality time series data collected from an Indian river.

III. Database

The HHT method is used for the analysis of four water quality parameters of Elunuthi Mangalam (E Mangalam) water quality station in Erode, belongs to Noyyal river in Tamil Nadu state in India. In India, there is a wide network of 390 water quality stations from different river basins under the Central Water Commission (CWC) which monitors 68 water quality parameters systematically. The suitability of river water for potable uses with regard to its physical and chemical quality has to be deciphered and defined on the basis of the some vital characteristics of water. Based on the assessment of water quality parameters from different stations, CWC identified water quality hotspots in India in the year 2011. E Mangalam station is one such station and it is the station in the network where the maximum number of water quality parameters fails to meet the permissible water quality criteria [14], i.e. water quality parameters exceeds the permissible limits. In this study, two physical parameters (temperature and Electrical Conductivity (EC)) and two chemical parameters (total hardness and magnesium (Mg)) are considered for the analysis. The monitoring is done in biweekly basis and the monthly data for the period November 2004-May 2012 (91 instances) collected from http://www.india-wris.nrsc.gov.in/ are used for the present study.

IV. Results and Discussion

The four water quality parameters are subjected to multiscale decomposition employing the CEEMDAN algorithm. Figure 1 shows the IMFs of four water quality parameters recorded at E Mangalam station. All the four water quality time series are found to be decomposable to six IMFs and a final residue, while the sixth mode is behaving more like a residue (oscillating with its own mean).

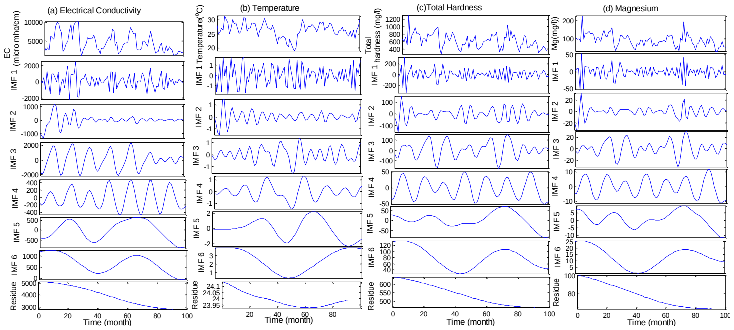

Figure 1. IMF components of four water quality parameters

The mean period of IMFs calculated by zero crossing method [10] and the correlation of IMFs with the respective series are presented in Table 1.

TABLE 1 MEAN PERIOD (T) OF IMFs AND THEIR CORRELATION (R) WITH ORIGINAL SERIES

IMF Number	Electrical Conductivity		Temperature		Total Hardness		Magnesium	
	R	T	R	T	R	T	R	T
IMF1	0.57	3.37	0.52	2.94	0.56	3.03	0.57	2.68
IMF2	0.54	6.07	0.46	6.50	0.44	6.07	0.41	6.07
IMF3	0.73	10.11	0.52	7.00	0.62	7.58	0.60	7.58
IMF4	0.61	13.00	0.54	13.00	0.47	13.00	0.49	13.00
IMF5	0.38	45.50	0.49	30.33	0.36	30.33	0.30	22.75
IMF6	0.45	45.50	0.45	45.50	0.45	45.50	0.41	45.50
Residue	0.43	91.00	0.28	91.00	0.39	91.00	0.42	91.00

It is noticed that the third IMF is showing the highest correlation with their respective series. The first three IMFs possess intra annual periodicity, the IMF4 is of near annual periodicity and the fifth mode is of inter annual periodicity. The final residue is denoting the inherent long term trend of the dataset and in this study, the trend of all the four water quality parameters is found to be decreasing.

The HSA of IMF of all water quality time series are performed and the instantaneous frequency trajectories are presented in Fig. 2.

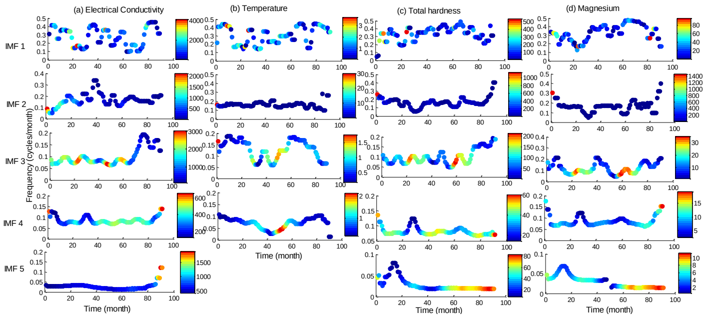

Figure 2. Instantaneous frequency trajectories of four water quality parameters of Noyyal river. The color bar represents the amplitude in respective units (EC- micro mhos/cm; Temperature-^{o}C; Total hardness- mg/l; Magnesium-mg/l)

From Fig. 2, it is noticed that high intermittency is present in IMF1. This is a typical characteristics of nonlinearity of the time series. Also a high frequency modulation is present in the low frequency part of the spectra, which show the non-stationarity of the data. Further it is noticed that the dominant frequency (where a concentration of highest amplitude occurs) is not constant, but varying with time. It is further noticed that the instantaneous frequency trajectories of different IMFs of the total hardness and magnesium time series are strikingly similar in terms of the time instants of peak, instant of occurrence of the highest amplitudes and overall non-stationarity (features of such as trend, seasonality) etc.

The integration of Hilbert spectra of IMFs lead to the marginal Hilbert spectrum, which offers a measure of total amplitude contribution from each frequency value. It represents the cumulated amplitude over the entire data span in a probabilistic sense. i.e., the amplitude at a frequency ω means there is a higher likelihood that an oscillation with such a frequency exists.

The marginal Hilbert Spectrum (MHS) of IMFs are obtained and the mean MHS of all the four water quality parameters are presented in Fig. 3.

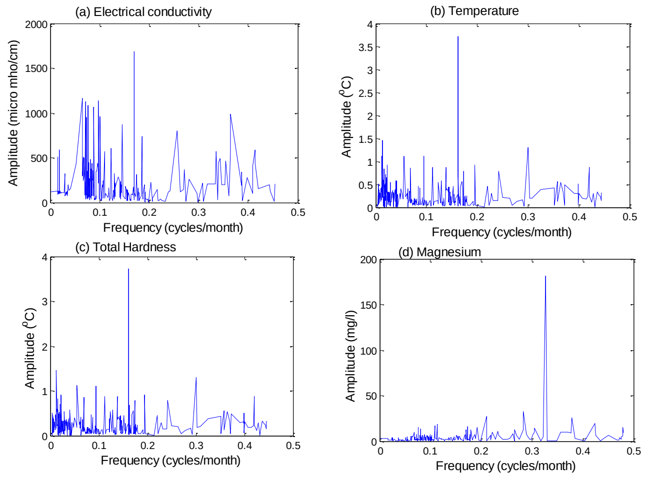

Figure 3. Mean marginal Hilbert spectrum of the four water quality parameters

The figure shows that the prominent peaks are in the frequency exceeding 0.18 in all the cases (i.e., high frequency range). This indicates that the likelihood of highest amplitude (refers more pollution) occurs in intra annual scale. This can be linked with anthropogenic impacts such as frequent disposals of pollutant load into the river.

The instantaneous amplitudes of IMFs of discharge and the four water quality parameters are subjected to a trend analysis by the modified Mann Kendall (MK) test [15] at 5 % significance level. The MK values are shown in Table 2. It is noted that the amplitudes of fifth mode (of inter annual periodicity) is showing a statistically significant increasing trend for the study period. However from Fig. 1 it is noted that the overall trend is reducing. Thus it can be inferred that the lower modes (high frequency) are responsible for such changes in the river system.

TABLE 2 MANN KENDALL VALUES OF INSTANTANEOUS AMPLITUDES OF WATER QUALITY PARAMETERS

IMF Number	EC	Temperature	Total hardness	Mg
IMF1	-1.429	**-2.069**	**-2.311**	-1.282
IMF2	**-5.062**	**-2.043**	-0.337	0.314
IMF3	**-5.392**	**2.039**	**-2.251**	0.082
IMF4	**2.938**	-0.301	1.445	**2.921**
IMF5	**5.177**	**10.131**	**8.423**	**4.775**

It is hoped that the information gained from this study may offer more realistic representation of the transport processes in the river, thus providing accurate results for future management and decision making. Further analysis can be made by coupling the multiscale decomposition with data-driven methods or linear models (such as auto regressive (AR)) to forecast the water quality parameters of prime concern and it is possible to eventually evaluate the risk associated with the existing water quality standards.

v. Conclusions

This paper presented an effective approach for analyzing water quality time series employing the multiscale decomposition by CEEMDAN algorithm and subsequent application of NHT. The method is demonstrated with four water quality parameter time series data collected from Noyyal river in India. The spectral representation clearly depicts the nonlinearity and non-stationarity of the datasets and the time varying behavior of dominant frequency. The similarity in the instantaneous frequency trajectories of magnesium and total hardness suggests that the primary contributor of the hardness of the river system is magnesium. The marginal Hilbert spectrum show that the prominent peaks of the different series are in the high frequency range, which may be attributed to the frequent pollutant disposal by anthropogenic interventions. The trend analysis of spectral amplitudes show that eventhough the IMF5 of inter annual periodicity is significantly increasing for all the parameters, as the overall trend is found to be decreasing, this may be because of the dominant effect of high frequency IMFs of intra annual or annual periodicities.

References

[1] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q Zheng, N.C. Yen, C.C. Tung, and H.H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” Proc.of Roy. Soc. London, Series A, vol. 454, pp. 903–995,1998

[2] N.E. Huang, and Z. Wu, “A review on Hilbert Huang Transform: Method and its applications to geophysical studies,” Rev. of Geophy, vol 46(2), doi: 10.1029/2007RG000228, 2008

[3] J. Rudi, R. Pabel, G. Jager, R. Koch, A. Kunoth, and H. Bogena,. “Multiscale analysis of hydrologic time series data using the Hilbert Huang Transform,” Vad. Zone J, vol. 9, pp. 925-942, 2010

[4] N. Massei, and M. Fournier, “Assessing the expression of large scale climatic fluctuations in the hydrologic variability of daily Seine river flow (France) between 1950-2008 using Hilbert Huang Transform,” J. Hydrol. vol. 448-449 , pp. 119-128, 2012

[5] Y.C. Chen, S.P. Kao, and H.W. Chiang , “Defining an estuary using the Hilbert-Huang transform,” Hydrol. Sci. J., vol. 58 (4), pp. 841–853, 2013

[6] A. Antico, G. Schlotthauer, and M.E. Torres. “Analysis of hydro-climatic variability and trends using a novel empirical mode decomposition: Application to Parana river basin,” J. of Geophys. Res.: Atm., doi: 10.1002/2013DJD020420, 2014

[7] A.R. Rao , and E.C. Hsu, Hilbert-Huang Transform Analysis of Hydrological and Environmental Time Series, Springer Verlag, Netherlands, 2008

[8] S. Franceschini, and C. W. Tsai, “Application of Hilbert-Huang Transform method for analyzing toxic concentrations in the Niagara river,” J. Hydrol. Engng, vol. 15(2), pp. 90-96, 2010

[9] M. Naixia, C. Cuixia, L. Weenbao, and W. Hui, “Hilbert-Huang Transform Based Time Series Analysis of Water Quality Parameters,” IEEE Fourth international symposium on Knowledge Acquisition and modeling, Sanya, China, pp.365-369, 2011

[10] Y. Huang, and F.G. Schmitt, “Time dependent intrinsic correlation analysis of temperature and dissolved oxygen time series using empirical mode decomposition,” J. Mar. Sys, vol. 130, pp. 90-100, 2014

[11] Z. Wu, and N.E. Huang, “Ensemble Empirical Mode Decomposition: A noise-assisted data analysis method,” Centre for Ocean-Land-Atmospheric Studies Tech. Rep. 193, Cent. for Ocean-Land-Atmos. Stud., Calverton, Md. 1-51 (ftp://grads.iges.org/pub/ctr/ctr_193.pdf)

[12] M.E. Torres, M.A. Colominas, G. Schlotthauer, and P. Fladrin, “A complete Ensemble Empirical Mode Decomposition with Adaptive Noise,” IEEE International conference on Acoustic Speech and Signal Processing, Prague 22-27 May 2011, pp. 4144-4147

[13] N.E. Huang, Z. Wu, S.R. Long, K.C. Arnold, K. Blank, and T.W. Liu “On instantaneous frequency,” Adv. in Adaptive Data Analy. vol. 01, pp. 177,doi: 10.1142/S1793536909000096, 2009

[14] CWC “Water Quality hotspots in India-A report by Central Water Commission,” 2011

[15] K. H. Hamed, and A.R. Rao, “A modified Mann-Kendall trend test for autocorrelated data, ” J. Hydrol, vol 204, pp. 182–196, 1998

About Author (s):

ADARSH S is working as assistant professor at TKM college of Engineering Kollam Kerala, India since 2004. He completed masters in Water Resources Engineering from IIT Bombay in 2009 and currently pursuing Ph.D at IIT Bombay. He published more than 15 papers in reputed international journals and many in conferences. His areas of research interests include the application of stochastic search algorithms for optimization of water resources and environmental systems, predictive modeling in hydrology and environmental engineering using data driven techniques, multiscale time series analysis etc.

M. JANGA REDDY, Ph.D. (IISc Bangalore India), is a faculty at Department of Civil Engineering, Indian Institute of Technology Bombay, India, working widely in the area of hydraulics and water resource engineering. He has published several technical papers in leading international journals on applications of evolutionary computation & soft computing techniques in water resources and environmental management.