

# Analysis of Water Parameters Using Daubechies Wavelet (Level 5) (Db5)

Kulwinder Singh Parmar, Rashmi Bhardwaj\*

Department of Mathematics, University School of Basic and Applied Sciences, Guru Gobind Singh Indraprastha University, Dwarka, Delhi, 110075, India

**Abstract** Statistical analysis of the water quality parameters COD (Chemical Oxygen Demand), BOD (Biochemical Oxygen Demand), DO (Dissolved Oxygen), WT (Water Temperature), AMM (Free Ammonia), TKN (Total Kjeldahl Nitrogen), TC (Total Coliform), FC (Fecal Coliform) and pH (Potential of Hydrogen) monitored at Hathni Kund Dam (Haryana) Yamuna River in India have been studied. It has been observed that water quality parameter have positive correlation between COD-BOD; AMM-TKN; WT-pH; TC-FC; and negatively correlation between COD-DO; BOD-DO; TKN-FC; DO-WT. Discrete wavelet analyses of water quality parameter using Daubechies wavelet at level 5 have been calculated. It has been observed that the values of signal data at five different levels for AMM, BOD, COD, DO, FC, pH, TC, TKN, WT varies between -1 to 1, -2 to 2, -20 to 20, -2 to 2, -5 to 5, -1 to 1, -2 to 2, -5 to 5, -10 to 10 respectively.

**Keywords** Wavelet Analysis, Water Pollutant, Statistical Analysis, Daubechies Wavelet

## 1. Introduction

Water is the basic requirement for life on earth. Many human activities pollute the river water. Urbanization, industrialization and globalization increased pollution level in river water. Yu, B. et al. (2010) [9] discussed the unsteady evolution of  $\text{NH}_3\text{-N}$  in the upper and middle reaches of Minjiang River through multi-scale analysis. Zainuddin, Z. and Pauline, O. (2011) [10] described the predictive capability of Wavelet Neural Networks (WNNs). He, Li et al. (2008) [5] studied wavelet-based multiresolution analysis technique (WMAT) for reducing noises induced by complex uncertainty. Hanbaya, D. et al. (2010) [4] studied switch circuits model based on wavelet decomposition and neural network. Kumar N. and Sinha D.K. (2010) [6] studied statistical regression analysis of underground drinking water. Vassilis Z. et al. (2001) [8] studied the time series analysis using statistical methods for monthly values of water quality parameters and the discharge. Can, Z. et al. (2005) [2] analyzed the characteristics of gravity waves (GWs) and seasonal variations of atmospheric waves over Istanbul using wavelet techniques. Zhang et al. (2008) [11] explained complex system dynamic (SD) model for water resources. Clonda, D. et al (2004) [3] discussed symmetric solution using complex Daubechies wavelets. Bhardwaj et al. (2011) [1] calculated water quality index (WQI) for 16 water quality parameters including 6 cations, 5 anions beside Dissolved Oxide (DO),

Biochemical Oxygen Demand (BOD), Total Dissolved Solids (TDS), Potential of Hydrogen (pH) and conductivity at the Hathni Kund Dam (Haryana, India). This paper deals with Statistical and Wavelet Analysis using Daubechies Wavelet at level 5 (Db 5) of water quality parameters.

## 2. Statistical Analysis of Water Quality Parameters

The monthly average values of last 10 years of water quality parameters COD (Chemical Oxygen Demand), BOD (Biochemical Oxygen Demand), DO (Dissolved Oxygen), WT (Water Temperature), AMM (Free Ammonia), TKN (Total Kjeldahl Nitrogen), TC (Total Coliform), FC (Fecal Coliform) and pH (Potential of Hydrogen) observed by Central Pollution Control Board (CPCB) at Hathni Kund Dam (Haryana) of Yamuna river (India) have been considered for study. In Statistical analysis mean, median, mode, minimum, maximum, range, standard deviation, kurtosis, skewness and coefficient of variation have been studied. Skewness gives the symmetry of data and coefficient of variation describes the relative measure of sample. Kurtosis ( $k$ ) refers the degree of flatness or peakedness and it also indicates impulsiveness of the signal. It defined as

$$k = \frac{1}{\sigma^4} \int_{-\infty}^{\infty} (x - \bar{x})^4 f(x) dx \quad (1)$$

where  $f(x)$  probability density function of the instantaneous amplitude  $x(t)$  at time  $t$ .  $\bar{x}$  is a mean value and  $\sigma$  is the standard deviation of  $x(t)$ .

The statistical analysis of each water quality parameters

\* Corresponding author:

rashmib22@gmail.com (Rashmi Bhardwaj)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2012 Scientific & Academic Publishing. All Rights Reserved

pH, COD, BOD, AMM, TKN, DO, WT, TC and FC have been discussed below as given in Table 1.

**pH:** Average, positional average and mode value of pH is 7.77, 7.72 and 7.7. These values are close to 7.7 thus data exhibits normal behavior. Standard deviation (SD) is 0.422, skewness approximates to 0 thus pH is symmetrical and values are very close to each other. Curve is platykurtic as Kurtosis is less than 3.

**COD:** Mean, median and mode value is 7.2, 6.0 and 5.0 respectively thus curve is not normal. Standard deviation value is high (5.6932807) thus values of COD are not close to each other. It is skewed (2.435) and curve is leptokurtic because Kurtosis is more than 3.

**BOD:** Mean and median value is approximately equal but mode value is different so curve is not normal. Standard deviation value suggests that data is spread out and curve is leptokurtic.

**AMM:** Average, median and mode value is approximately equal thus data exhibits normal behavior. Standard deviation (0.2879) suggests that data is close to each other. Curve is platykurtic.

**TKN:** Mean and median value is approximately equal. Standard deviation (0.83) suggests that sample data is close to each other. Skewness value is approximately 0 thus curve is symmetrical and platykurtic.

**DO:** Average, median and mode value is equal thus data behaves normally. Standard deviation value is 1.344 curve is symmetric and platykurtic.

**WT:** Mean, median and mode value is approximates to 21 thus data exhibits normal characteristic. Curve is skewed and platykurtic.

**TC:** Average, median and mode value is different. Standard deviation value explains that sample data is spread. Curve is leptokurtic.

**FC:** Average, median and mode value is different. Standard deviation value explains that sample data is spread. Curve is leptokurtic. Curve is neither symmetrical nor skewed.

### 3. Correlation Coefficient between Water Quality Parameter

The correlation coefficient is defined as

$$r_{xy} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{(n-1) \delta_x \delta_y} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \quad (2)$$

where  $\bar{x}$  and  $\bar{y}$  are the sample means of X and Y, and  $\delta_x$  and  $\delta_y$  are the sample standard deviations of X and Y.

Table-2, gives the value of correlation coefficient between pairs of water quality parameters. It has been observed that COD is positively correlated with BOD and negatively correlated with DO. Positive correlation means persistent behavior with the variables. BOD is negatively correlated with DO. AMM is positively correlated with TKN. TKN is negatively correlated with TC. DO is negatively correlated with WT. WT is positively correlated with pH and TC is positive correlated with FC. Figure 1, plots the graphical representation of correlation coefficient of water quality parameters with other parameters.

**Table 1.** Statistical Analysis of Yamuna River at Hathni Kund Dam

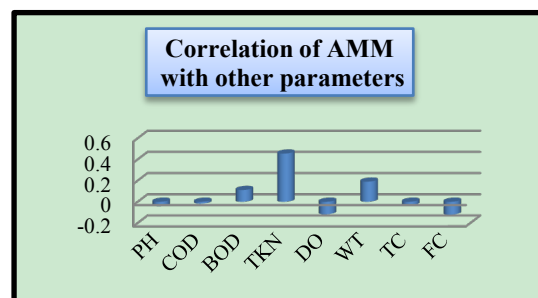
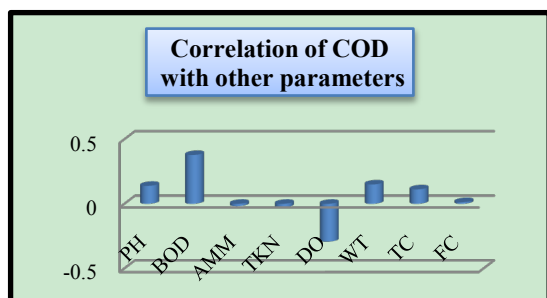
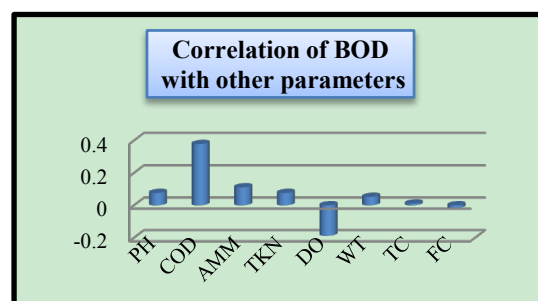
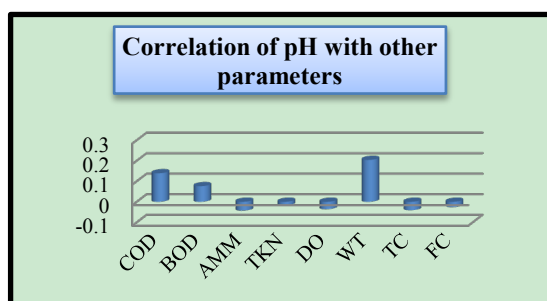
Parameters	pH	COD	BOD	AMM	TKN	DO	WT	TC	FC
Mean	7.77	7.20	1.28	0.20	1.02	9.37	20.51	6,40,808.74	18,177.11
Median	7.72	6.00	1.00	0.09	0.84	9.40	21.75	65,000.00	1,105.00
Mode	7.7	5	1	0	0.27	9.6	22	21000	40
Minimum	6.83	1.00	1.00	0.00	0.00	6.60	10.00	100.00	0.00
Maximum	9.00	36.00	4.00	1.33	6.16	13.00	29.00	3,30,00,000.00	6,90,000.00
Range	2.17	35.00	3.00	1.33	6.16	6.40	19.00	3,29,99,900.00	6,90,000.00
SD	0.4219906	5.6932807	0.5792272	0.2879544	0.8304988	1.343981	4.8874171	3116814.156	69578.13012
Kurtosis	0.4008068	8.0001476	5.2042797	2.4510663	12.432353	-0.2292778	-0.9596681	100.6365702	74.70462933
Skewness	0.423504	2.4348628	2.2676231	1.7049319	2.7330303	0.2411506	-0.4004147	9.717438876	8.045445233
Coeff. Of variation	0.0543277	0.7907334	0.4542959	1.4132733	0.8128863	0.1434217	0.2383331	4.863875855	3.827788714

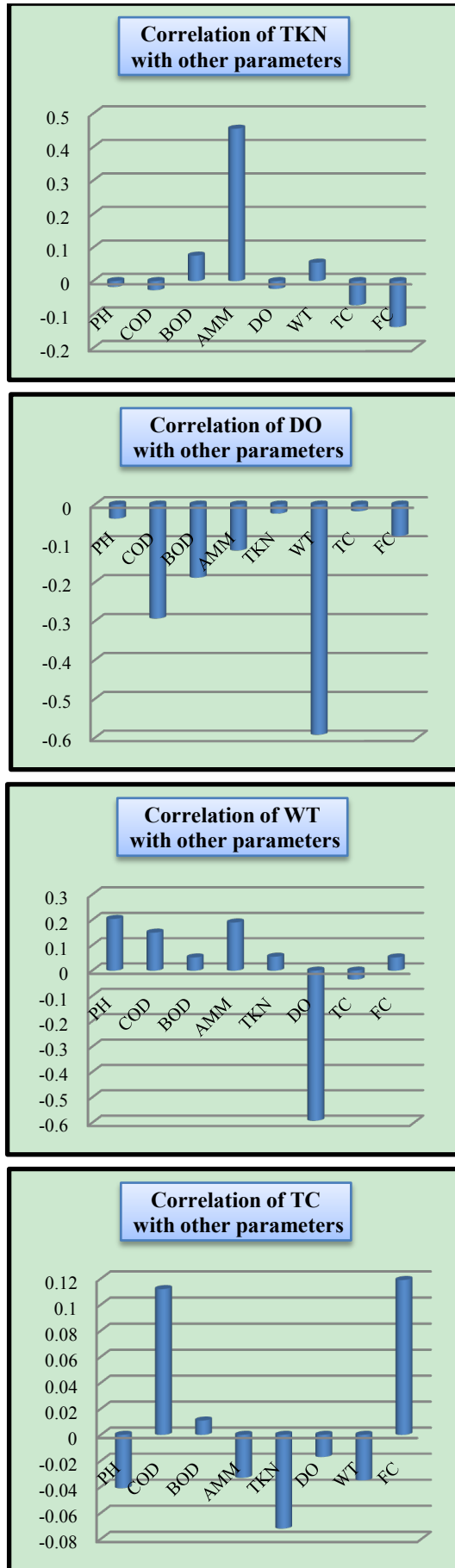
**Table 2.** Correlation between water quality parameters at Hathni Kund (Haryana) Correlations

		PH	COD	BOD	AMM	TKN	DO	WT	TC	FC
PH	Pearson Correlation	1	.139	.076	-.043	-.017	-.036	.204*	-.041	-.026
	Sig. (2-tailed)		.130	.411	.639	.853	.696	.025	.656	.778
	N	120	120	120	120	120	120	120	120	120
COD	Pearson Correlation	.139	1	.378**	-.018	-.027	-.293**	.150	.112	.013
	Sig. (2-tailed)	.130		.000	.848	.766	.001	.102	.223	.892
	N	120	120	120	120	120	120	120	120	120
BOD	Pearson Correlation	.076	.378**	1	.111	.076	-.188*	.052	.011	-.018
	Sig. (2-tailed)	.411	.000		.229	.408	.040	.571	.901	.841
	N	120	120	120	120	120	120	120	120	120
AMM	Pearson Correlation	-.043	-.018	.111	1	.455**	-.118	.190*	-.033	-.121
	Sig. (2-tailed)	.639	.848	.229		.000	.198	.037	.720	.189
	N	120	120	120	120	120	120	120	120	120
TKN	Pearson Correlation	-.017	-.027	.076	.455**	1	-.023	.055	-.072	-.137
	Sig. (2-tailed)	.853	.766	.408	.000		.807	.553	.438	.137
	N	120	120	120	120	120	120	120	120	120
DO	Pearson Correlation	-.036	-.293**	-.188*	-.118	-.023	1	-.592**	-.017	-.081
	Sig. (2-tailed)	.696	.001	.040	.198	.807		.000	.854	.381
	N	120	120	120	120	120	120	120	120	120
WT	Pearson Correlation	.204*	.150	.052	.190*	.055	-.592**	1	-.035	.052
	Sig. (2-tailed)	.025	.102	.571	.037	.553	.000		.706	.571
	N	120	120	120	120	120	120	120	120	120
TC	Pearson Correlation	-.041	.112	.011	-.033	-.072	-.017	-.035	1	.119
	Sig. (2-tailed)	.656	.223	.901	.720	.438	.854	.706		.195
	N	120	120	120	120	120	120	120	120	120
FC	Pearson Correlation	-.026	.013	-.018	-.121	-.137	-.081	.052	.119	1
	Sig. (2-tailed)	.778	.892	.841	.189	.137	.381	.571	.195	
	N	120	120	120	120	120	120	120	120	120

\*. Correlation is significant at the 0.05 level (2-tailed).

\*\*. Correlation is significant at the 0.01 level (2-tailed).





**Figure 1.** Pearson correlation coefficient of (a) pH (b) COD (c) BOD (d) AMM (e) TKN (f) DO (g) WT (h) TC and (i) FC with other parameters.

#### 4. Discrete Wavelet Analysis of Water Quality Parameters

A wavelet is a function  $\psi \in L^2(\mathbb{R})$  which satisfies the condition

$$C_\psi \equiv \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \quad (3)$$

where  $\hat{\psi}(\omega)$  is the Fourier transform of  $\psi(t)$

It may be observed that the wavelet transform is a prism which exhibits properties of signal such as points of abrupt changes, seasonality or periodicity. The wavelet transform is a function of 'a' and 'b' where

a = the scale of frequency

b = spatial position (translation) or time

The plane defined by the variables (a, b) is called the scale-space or time frequency plane. A continuous representation of a function of two continuous parameters a, b can be converted into a discrete one by assuming that a and b take only integral values.

$$\psi_{a,b}(t) = a^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right), \text{ where } a > 0 \quad (4)$$

$T_\psi f(a,b)$  is called the Wavelet Transform of  $f(t)$  and

$$T_\psi f(a,b) = a^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt = \langle f, \psi_{a,b} \rangle$$

= inner product or dot product of  $f$  and  $\psi_{a,b}$

If  $\psi \in L^2(\mathbb{R})$ , then  $\psi_{a,b}(t) \in L^2(\mathbb{R})$  for all a, b.

$$\|\psi_{a,b}(t)\|^2 = |a|^{-1} \int_{-\infty}^{\infty} \left| \psi\left(\frac{t-b}{a}\right) \right|^2 dt = \int_{-\infty}^{\infty} |\psi(x)|^2 dx = \|\psi\|^2 \quad (5)$$

The Fourier transform is given by

$$\hat{\psi}_{a,b}(\omega) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-i\omega t} \psi\left(\frac{t-b}{a}\right) dt = |a|^{-\frac{1}{2}} e^{-i\omega b} \hat{\psi}(a\omega) \quad (6)$$

One dimensional discrete wavelet analysis of monthly average value for last ten years of water quality parameter

such as pH, BOD, COD, DO, AMM, TKN, WT, TC and FC for Yamuna river at Hathni Kund Dam, Haryana using ten years values. Daubechies wavelet at level 5 (Db 5) contains seven parts  $s$ ,  $a_5$ ,  $d_1$ ,  $d_2$ ,  $d_3$ ,  $d_4$  and  $d_5$ . The first part ' $s$ ' represents the signal or raw data and the second part ' $a_5$ ' corresponds to the amplitude of the signal. The last parts  $d_1$ ,  $d_2$ ,  $d_3$ ,  $d_4$  and  $d_5$  represents details of the signal or raw data at five different levels.

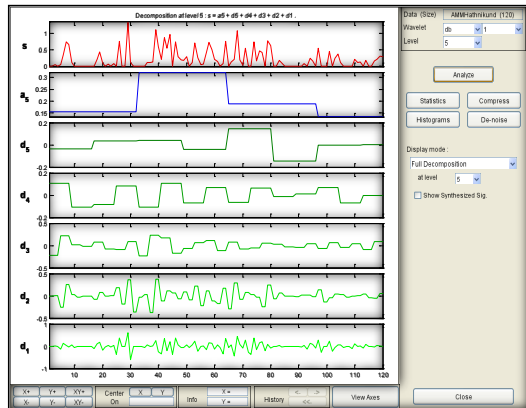


Figure 2(a). 1D Discrete wavelet for AMM

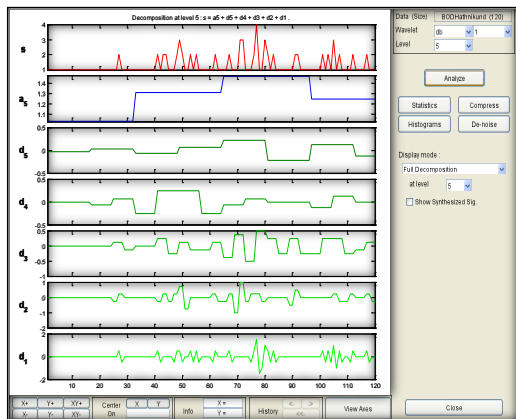


Figure 2(b). 1D Discrete wavelet for BOD

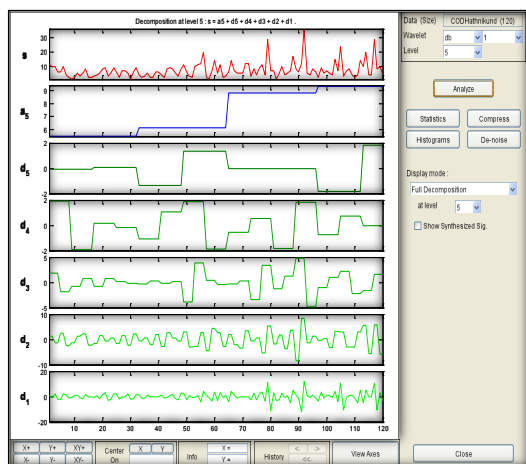


Figure 2(c). 1D Discrete wavelet for COD

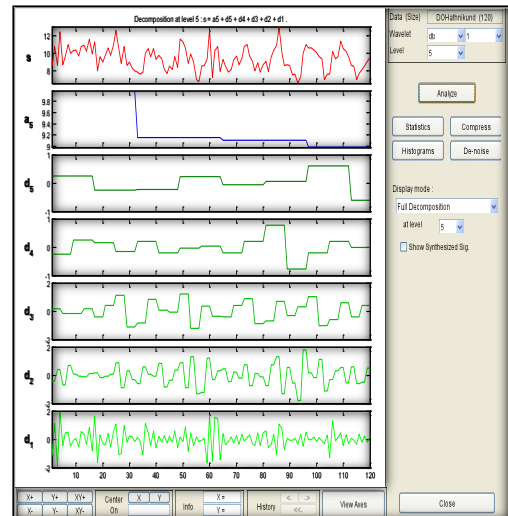


Figure 2(d). 1D Discrete wavelet for DO

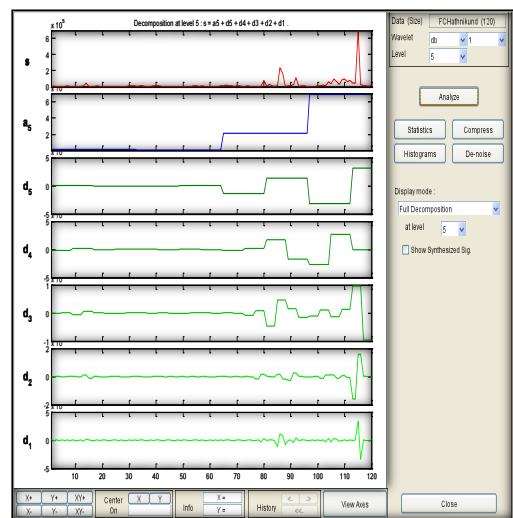


Figure 2(e). 1D Discrete wavelet for FC

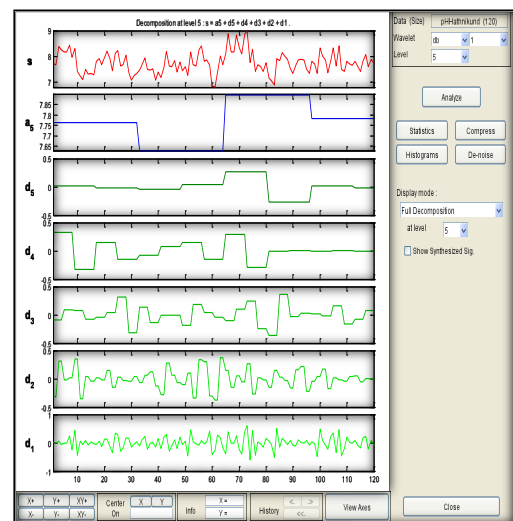


Figure 2(f). 1D Discrete wavelet for pH

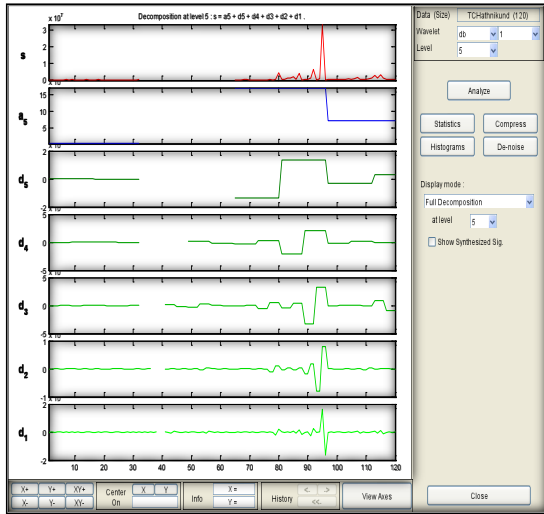


Figure 2(g). 1D Discrete wavelet for TC

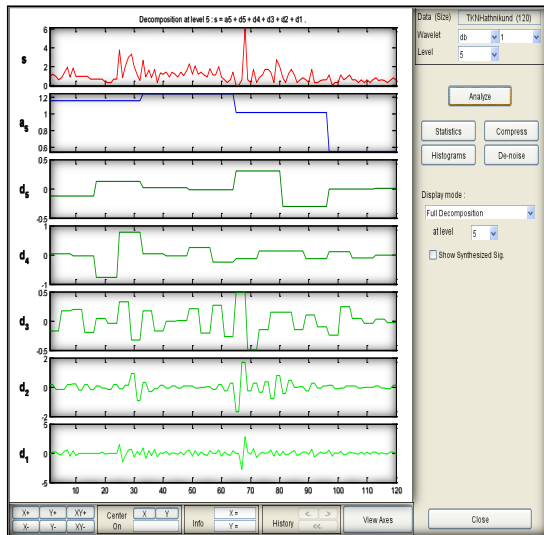


Figure 2(h). 1D Discrete wavelet for TKN

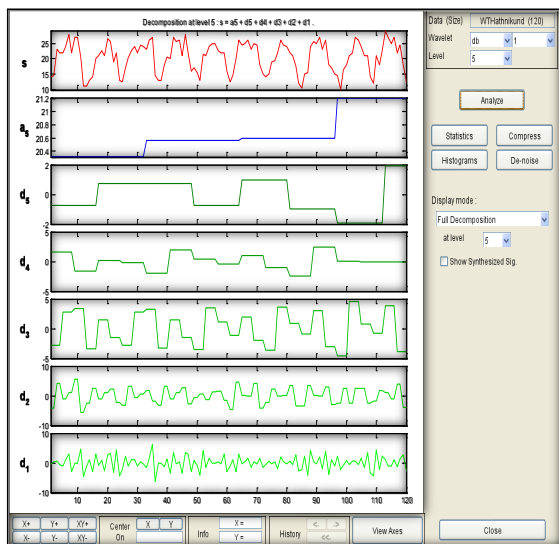


Figure 2(i). 1D Discrete wavelet for WT

Figure 2, gives the details of 1D Daubechies Wavelet analysis at level 5 (Db 5) for each water quality parameter. It

has been observed that:

**AMM:** The first part of the figure 2(a) shows that maximum value of AMM is 1.33 and the corresponding amplitude ( $a_5$ ) varies from 0.15 to 0.35. The values of  $d_5$ ,  $d_4$ ,  $d_3$ ,  $d_2$  and  $d_1$  vary from -0.2 to 0.2, -0.2 to 0.2, -0.5 to 0.5, -0.5 to 0.5 and -1 to 1 respectively.

**BOD:** Figure 2(b) explains the BOD parameter's signal and maximum value is 4 and amplitude ( $a_5$ ) of this signal varies from 1.0 to 1.5. The other values of  $d_5$ ,  $d_4$ ,  $d_3$ ,  $d_2$  and  $d_1$  varies from -0.5 to 0.5, -0.5 to 0.5, -1 to 0.5, -2 to 1 and -2 to 2 respectively.

**COD:** Figure 2(c) explains the COD parameter's signal and maximum value is 36 and amplitude ( $a_5$ ) varies from 5.5 to 9.5. The other values of  $d_5$ ,  $d_4$ ,  $d_3$ ,  $d_2$  and  $d_1$  varies from -2 to 2, -2 to 2, -5 to 5, -10 to 10 and -20 to 20 respectively.

**DO:** Figure 2(d) describes the DO parameter's signal and maximum value is 13 and amplitude ( $a_5$ ) varies from 9.0 to 9.9. The other values of  $d_5$ ,  $d_4$ ,  $d_3$ ,  $d_2$  and  $d_1$  vary from -1 to 1, -1 to 1, -2 to 2, -2 to 2 and -2 to 2 respectively.

**FC:** Figure 2(e) explains the FC parameter's signal and maximum value is  $6 \times 10^5$  and amplitude ( $a_5$ ) varies from 0 to 7. The other values of  $d_5$ ,  $d_4$ ,  $d_3$ ,  $d_2$  and  $d_1$  vary from -5 to 5, -5 to 5, -1 to 1, -2 to 2 and -5 to 5 respectively.

**pH:** Figure 2(f) depicts the pH parameter's signal and maximum value is 9 and amplitude ( $a_5$ ) varies from 7.64 to 7.85. The other values of  $d_5$ ,  $d_4$ ,  $d_3$ ,  $d_2$  and  $d_1$  varies from -0.5 to 0.5, -0.5 to 0.5, -0.5 to 0.5, -0.5 to 0.5 and -1 to 1 respectively.

**TC:** First part of the Figure 2(g) explains the TC parameter's signal and maximum value as  $3 \times 10^7$  and amplitude ( $a_5$ ) varies from 0 to 17. The other values of  $d_5$ ,  $d_4$ ,  $d_3$ ,  $d_2$  and  $d_1$  vary from -2 to 2, -5 to 5, -5 to 5, -1 to 1 and -2 to 2 respectively.

**TKN:** Figure 2(h) explains the TKN parameter's signal and gives maximum value is 6.16 and amplitude ( $a_5$ ) varies from 0.6 to 1.2. The other values of  $d_5$ ,  $d_4$ ,  $d_3$ ,  $d_2$  and  $d_1$  varies from -0.5 to 0.5, -1 to 1, -0.5 to 0.5, -2 to 2 and -5 to 5 respectively.

**WT:** Figure 2(i) explains the WT parameter's signal and maximum value is 29 and amplitude ( $a_5$ ) varies from 20.4 to 21.1. The other values of  $d_5$ ,  $d_4$ ,  $d_3$ ,  $d_2$  and  $d_1$  vary from -2 to 2, -5 to 5, -5 to 5, -10 to 10 and -10 to 10 respectively.

## 5. Conclusions

Statistical and wavelet analysis of water quality parameters monitored at the Hathni Kund Dam (Haryana) of Yamuna River (India) have been discussed. It has been observed that curve is platykurtic for pH, AMM, TKN, DO, WT; leptokurtic for COD, BOD, TC, FC and normal for pH, AMM, DO, WT. There is positive correlation between COD-BOD, AMM-TKN, WT-pH, TC-FC and negative correlation between COD-DO, BOD-DO, TKN-FC, DO-WT. 1-D Discrete Daubechies Wavelet analysis explains that the parameters AMM, BOD, COD, DO, FC, pH, TC, TKN and WT have the maximum value 1.33, 4.0, 36, 13,

$6 \times 10^5$ , 9.0,  $3 \times 10^7$ , 6.16 and 35 and amplitude ( $a_5$ ) varies between 0.15 to 0.35, 1.0 to 1.5, 5.5 to 9.5, 9.0 to 9.9, 0 to 7, 7.64 to 7.85, 0 to 17, 0.6 to 1.2 and 20.4 to 21.1 respectively. The values of  $d_5$ ,  $d_4$ ,  $d_3$ ,  $d_2$  and  $d_1$  ranges between -1 to 1 for AMM, pH; -2 to 2 for BOD, DO, TC; -5 to 5 for FC, TKN; -10 to 10 for WT and -20 to 20 for COD. It has been concluded that water quality parameters cross the prescribed limit of WHO/EPA.

## ACKNOWLEDGEMENTS

Authors are thankful to Guru Gobind Singh Indraprastha University for providing research facilities and financial support. Authors also thankful to Central Pollution Control Board (CPCB) for providing data for research work.

## REFERENCES

- [1] Bhardwaj, R., Parmar, K.S., Chuha, P., Minhas, P. and Sahota, H.S. (2011) "Seasonal variation of physico-chemical parameters and water quality indexing of Harike Lake" *Indian j. Env. Prot.*, 31(6), 482-486.
- [2] Can, Z.; Aslan, Z., Oguz, O. and Siddiqi, A.H. (2005) "Wavelet Transform of Metrological Parameter and Gravity Waves" *Annales Geophysicae*, 23, 659-663.
- [3] Clonda, D., Lina, J. and Goulard, B. (2004) "Complex Daubechies wavelets: properties And statistical image modeling" *Signal Processing*, 84 (1), 1-23.
- [4] Hanbaya, D., Turkoglua, I., and Demirb, Y. (2010) "Modeling switched circuits based on wavelet decomposition and neural networks" *Journal of the Franklin Institute*, 347(3), 607-617.
- [5] He, L., Huang, G., Zeng, G. and Lu, H. (2008) "Wavelet-based multiresolution analysis for data cleaning and its application to water quality management systems" *Expert Systems with Applications*, 35 (3), 1301-1310.
- [6] Kumar, N. and Sinha, D.K. (2010) "Drinking water quality management through Correlation studies among various physico-chemical parameters: A case study" *I.J. Env. Sci.*, 1(2), 253-259.
- [7] Shukla, J.B., Misra, A.K. and Chandra, P. (2008) "Mathematical Modeling and analysis of the dissolved oxygen in eutrophied water bodies affected by organic pollutant" *Non-linear Analysis: Real world Applications*, 9, 1851-1865.
- [8] Vassilis, Z., Papamichail, D.M. and Mitsiou (2001) "Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece" *Hydrology and Earth System Sciences*, 5(4), 679-691.
- [9] Yu, B., Huang, C.M., Liu, Z., Zhang, B. and Qin, F. (2010) "Wavelet Analysis on  $\text{NH}_3$ -N Pollution Index Changes of the Middle-Upper Minjiang River "during Last 6 Years" International Society for Environmental Information Sciences 2010 Annual Conference (ISEIS) *Procedia Environmental Sciences*, 2, 9-14.
- [10] Zainuddin, Z. and Pauline, O. (2011) "Modified wavelet neural network in function approximation and its application in prediction of time-series pollution data" *Applied Soft Computing*, 11 (8), 4866-4874.
- [11] Zhang, X.H., Zhang, H.W., Chen, B. and Zhao, X.H (2008) "Water resources planning based on complex system dynamics: A case study of Tianjin city" *Commun. Nonlinear Sci. Numer. Simul.*, 13(10), 2328-2336.