Water Science and Engineering, Dec. 2008, Vol. 1, No. 4, 82-88
ISSN 1674-2370, *http://kkb.hhu.edu.cn*, e-mail: *wse@hhu.edu.cn*

R/S method for evaluation of pollutant time series in environmental quality assessment

Bu Quanmin[1, 2], Bi Jun*[1, 2, 3], Yuan Zengwei[1, 2, 3], Huang Lei[1, 2]

1. State Key Laboratory of Pollution Control and Resources Reuse, Nanjing 210093, P. R. China
2. School of Environment, Nanjing University, Nanjing 210093, P. R. China
3. Center for Environmental Management and Policy, Nanjing University and Jiangsu Environmental Protection Bureau, Nanjing 210093, P. R. China

Abstract: The significance of the fluctuation and randomness of the time series of each pollutant in environmental quality assessment is described for the first time in this paper. A comparative study was made of three different computing methods: the same starting point method, the striding averaging method, and the stagger phase averaging method. All of them can be used to calculate the Hurst index, which quantifies fluctuation and randomness. This study used real water quality data from Shazhu monitoring station on Taihu Lake in Wuxi, Jiangsu Province. The results show that, of the three methods, the stagger phase averaging method is best for calculating the Hurst index of a pollutant time series from the perspective of statistical regularity.

Key words: *environmental quality assessment*; *time series*; *fractal dimension*; *R/S statistical method*; *Hurst index*

DOI: 10.3882/j.issn.1674-2370.2008.04.009

1 Introduction

After the evaluation index for environmental quality assessment is determined, the key task is to collect field data of the concentration of each target pollutant at a certain measuring point at different time. Data that indicate the variation of the concentration of a particular pollutant x with time t can be placed in a sequence (called a time series) as follows:

$$x(t_0), x(t_1), x(t_2), \cdots, x(t_n) \tag{1}$$

General testing of environmental quality usually has some regularity: data have usually been collected on either a daily or a weekly basis. Therefore, a precondition for the following discussion is

$$t_1 - t_0 = t_2 - t_1 = t_3 - t_2 = \cdots = t_n - t_{n\text{-}1} \tag{2}$$

Currently, the data x_i from a certain time period are used in almost all environmental quality assessments to calculate the arithmetic average value, which is considered the basic data for environmental quality assessment and calculation. The randomness of the time series is not taken into account. The importance of the average value of the concentration of each

This work was supported by the Eleventh Five-Year Key Technology R and D Program, China (Grant No. 2006BAC02A15), the Colleges and Universities in Jiangsu Province Natural Science-Based Research Projects (Grant No. 2006BAC02A15), the Jiangsu Province Post-Doctoral Fund Projects (Grant No. 0801006C), and the China Post-Doctoral Science Foundation (Grant No. 20080441032).
*Corresponding author (e-mail: *jbi@nju.edu.cn*)
Received Sep. 28, 2008; accepted Nov. 22, 2008

pollutant, which is the major tool for estimating the degree of membership and weight, should be acknowledged. Yet some valuable information is certainly lost with a disorderly time series.

There are reasons for fluctuations of time series. For example, the concentration of a certain pollutant in a lake is related to the water storage, water inflow and water output of the lake, as well as the rainfall and the amount of evaporation. These are all natural factors. There are also artificial factors, such as the irregular pollutant discharge from all the factories around the lake. The latter is a major cause of fluctuation. The amount of discharge from factories plays a decisive role in the average concentration $\overline{x}$ of a pollutant; the irregularity of the discharge affects the randomness of the concentration of pollutants. Therefore, two major tasks in environmental monitoring and environmental quality assessment are measuring the average value of the concentration of pollutants and understanding the degree of complexity of variation of the time series of pollutant concentration. These tasks are interdependent.

2 Fractal dimension and Hurst index

In 1982, fractal science came into being in the mathematical field. It is the science that Mandelbrot established, which specializes in randomness, irregularity and at the same time similarity of variation (Weffimuny and Pehumadu 2003). Using a certain physical quantity that varies continuously with time as an example, the characteristics of its variation are “continuous and nondifferentiable” (Jaw and Chen 1999). Fractal science studies neither the global average of the quantity, nor the trend of discretizing continuous physical quantities, nor the trend of fitting smooth curves. It studies the degree of complexity of random variation of the physical quantity, and uses the fractal dimension D as a quantitative index of this degree. In fractal science, the correlation dimension application has been adopted to compute the fractional dimension of the time series (Anh et al. 1997). At the same time, as Mandelbrot pointed out, in early years, the Hurst index H can be obtained by analyzing time series using the R/S analytical method, which can describe the degree of complexity of the variation of the physical quantity equally well (Kouzani et al. 1999). The relationship between D and H is $D = 2 - H$ (Castillo and Melin 2001). Professor Wu Zuoyong of Sichuan University has used this method to evaluate the environmental quality of air (Hambly and Jones 2003). Currently, no report on this method has been applied in the evaluation of water quality in China or other countries.

This study applied this method to the evaluation of water quality and focused on the possible divergences in processing methods when calculating H with the R/S method.

3 R/S analytical method

The basic idea of the R/S analytical method is the following:

For the time series of a certain physical quantity $x_1, x_2, x_3, \cdots, x_N$ at time $t_1, t_2, t_3, \cdots, t_N$, with a regular time interval, the time series span is

$$\Delta t = t_N - t_1 \tag{3}$$

Within the time period Δt, the average value of x_i is

$$< x >_N = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{4}$$

At time t_i, the accumulative deviation of the physical quantity x_i against the average value $< x >_N$ is

$$X(t_j, N) = \sum_{i=1}^{j}\left(x_j - < x >_N\right) \tag{5}$$

$X\left(t_i, N\right)$ varies with two parameters (Liebovitch 1997). The parameter *N* indicates the number of the physical quantity x_i in the time series. *N* is regarded as a variable, which changes with the selected statistical time during a certain time span. The value of parameter t_i is selected when the value of *N* is invariant. Stability of *N* is the precondition of each calculation.

The difference between the maximum $X\left(t, N\right)$ and the minimum $X\left(t, N\right)$ is defined as the range and referred to as *R*:

$$R\left(t_N - t_1\right) = R\left(\Delta t\right) = \max\left[\underset{t_1 \leqslant t \leqslant t_N}{X}\left(t, N\right)\right] - \min\left[\underset{t_1 \leqslant t \leqslant t_N}{X}\left(t, N\right)\right] \tag{6}$$

The value of $R\left(N\right)$ varies with *N*. As the value of *N* grows, so does the value of $R\left(N\right)$ (Ni and Liu 2005; Shen and Chen 2005). Yet the $R\left(N\right)$-N curve lacks regularity, and flat roofs even appear (Zhang et al. 2005; Ostrosi et al. 2003). In order to describe the regularity, Hurst made use of the standard deviation *S* of the value $X\left(t\right)$:

$$S = \left[\frac{1}{\Delta t}\sum_{i=1}^{N}\left(x_i - < x >_N\right)^2\right]^{1/2} \tag{7}$$

and defined the nondimensional ratio *R*/*S* as follows:

$$\frac{R}{S} = \frac{\max\left[\underset{t_1 \leqslant t \leqslant t_N}{X}\left(t, N\right)\right] - \min\left[\underset{t_1 \leqslant t \leqslant t_N}{X}\left(t, N\right)\right]}{\left[\frac{1}{\Delta t}\sum_{i=1}^{N}\left(x_i - < x >_N\right)^2\right]^{1/2}} \tag{8}$$

Hurst described ln(*R*/*S*) with ln*N* and found that the points in the data group almost lie on a straight line with perfect regularity (Wang et al. 2005; Shao et al. 2006). The slope of the line is called the Hurst index *H*. The relationship between *R*/*S* and *N* is R/S-N^H (Ikeda et al. 1999).

4 Three possible algorithms of R/S analytical method

There are three possible ways to compute $\left((R/S)_i, N_i\right)$ in order to calculate the Hurst index *H*:

(1) The same starting point method: When $N = 3$, $\left(x_1, x_2, x_3\right)$ is selected and $\left(R/S\right)_3$ is calculated; and when $N = 4$, $\left(x_1, x_2, x_3, x_4\right)$ is selected and $\left(R/S\right)_4$ is calculated.

(2) The striding averaging method: When $N = 3$, $\left(x_1, x_2, x_3\right)$, $\left(x_4, x_5, x_6\right)$, and $\left(x_7, x_8, x_9\right)$ are selected, and the corresponding *R*/*S*, whose mean value is $\left(R/S\right)_3$ is calculated; and when $N = 4$, $\left(x_1, x_2, x_3, x_4\right)$, $\left(x_5, x_6, x_7, x_8\right)$, and $\left(x_9, x_{10}, x_{11}, x_{12}\right)$ are selected, R/S is calculated and the value is averaged. If the statistical time is one year, and the real measurement is on a weekly basis, then there will be 52 numerical values and 13 groups of $\left(R/S\right)_4$ values

according to the *R*/*S* value.

(3) The stagger phase averaging method: When $N=3$, $\left(x_1, x_2, x_3\right)$, $\left(x_2, x_3, x_4\right)$, and $\left(x_3, x_4, x_5\right)$ are selected, thereby determining *R*/*S*, whose mean value is regarded as $\left(R/S\right)_3$. If there are *n* values, *n*-*N*+1 groups of data will come into being.

5 Demonstrations

To explore Hurst index calculation, three different ways of calculating the Hurst index have been studied. The processing program was based on the weekly water quality data from Shazhu monitoring station on Taihu Lake in Wuxi, Jiangsu Province from December 29, 2004 to June 25, 2007, which were released by the State Environmental Protection Administration (Table 1).

Table 1 Weekly water quality data from Shazhu monitoring station on Taihu Lake in Wuxi, Jiangsu Province

Week number	pH	DO (mg/L)	COD_{Mn} (mg/L)	NH_3-N (mg/L)	Week number	pH	DO (mg/L)	COD_{Mn} (mg/L)	NH_3-N (mg/L)	Week number	pH	DO (mg/L)	COD_{Mn} (mg/L)	NH_3-N (mg/L)
1	7.90	10.00	4.7	0.50	45	7.91	9.82	5.1	0.13	89	7.93	8.86	4.2	0.15
2	7.92	10.00	4.5	0.51	46	7.89	9.50	3.8	0.07	90	8.02	9.13	5.3	0.13
3	7.94	10.00	4.4	0.42	47	7.80	9.34	4.6	0.08	91	7.98	9.11	4.9	0.13
4	7.91	9.92	4.3	0.42	48	7.85	8.95	4.1	0.10	92	7.97	8.76	4.9	0.14
5	7.99	10.70	4.3	0.39	49	7.88	8.88	4.2	0.10	93	7.84	8.54	4.8	0.14
6	8.04	10.90	4.0	0.40	50	7.87	8.81	4.4	0.13	94	7.98	8.91	4.9	0.12
7	8.03	10.80	4.5	0.55	51	7.85	8.76	4.4	0.12	95	7.87	8.48	5.3	0.14
8	7.97	10.10	4.8	0.55	52	7.87	8.54	4.4	0.24	96	7.88	8.75	5.5	0.16
9	7.99	10.10	5.1	0.51	53	7.83	8.58	4.5	0.19	97	7.86	8.50	5.5	0.15
10	7.96	10.00	5.2	0.65	54	7.85	8.80	4.3	0.36	98	7.87	8.51	5.5	0.16
11	7.98	9.95	5.0	0.56	55	7.90	9.89	4.2	0.53	99	7.93	8.75	5.5	0.13
12	7.85	9.81	4.9	0.45	56	7.98	9.96	4.1	0.45	100	7.91	8.34	5.4	0.16
13	7.81	9.91	4.8	0.49	57	7.96	9.93	3.7	0.67	101	7.95	8.82	5.3	0.12
14	7.76	9.86	4.5	0.80	58	8.01	9.91	5.2	0.68	102	7.95	8.20	5.4	0.19
15	7.89	9.83	4.1	0.80	59	7.98	9.88	4.9	0.78	103	7.87	8.81	5.5	0.43
16	7.93	9.75	4.9	0.80	60	7.94	9.84	4.7	0.88	104	7.85	9.61	5.7	0.78
17	7.95	9.16	5.1	0.88	61	7.97	9.82	4.8	1.19	105	7.86	9.31	5.4	0.32
18	7.93	9.81	5.3	0.80	62	7.99	9.81	4.9	1.49	106	8.00	8.97	5.5	0.46
19	7.93	9.78	4.8	0.71	63	7.95	9.02	5.4	1.39	107	7.84	8.98	4.8	0.50
20	7.89	9.84	5.1	0.41	64	7.88	9.23	5.9	1.19	108	8.04	8.73	5.1	0.65
21	7.92	9.96	4.9	0.48	65	7.85	9.07	5.2	1.33	109	7.95	9.08	4.5	0.57
22	7.94	9.94	4.9	0.47	66	7.90	9.30	4.7	0.97	110	7.95	8.58	3.9	0.45
23	7.95	9.91	4.9	0.45	67	7.89	8.76	4.4	0.23	111	7.95	8.81	5.0	0.54
24	7.97	9.97	4.7	0.47	68	7.92	8.68	4.7	0.23	112	7.94	9.32	5.5	0.65
25	7.92	9.86	4.9	0.44	69	7.93	9.22	4.4	0.14	113	7.85	8.96	5.2	0.54
26	7.89	9.92	4.9	0.45	70	7.93	9.07	4.4	0.29	114	7.87	9.74	5.8	0.85
27	7.92	9.89	4.7	0.54	71	7.93	8.96	4.2	0.20	115	7.87	9.02	5.6	0.56
28	7.97	9.92	4.9	0.55	72	7.96	9.11	5.7	0.32	116	7.88	9.07	5.7	0.68
29	7.93	9.75	5.4	0.63	73	7.90	9.73	5.7	0.13	117	7.87	9.47	5.6	0.74
30	7.87	9.87	4.8	0.58	74	7.98	9.57	6.6	0.15	118	7.91	8.65	5.5	0.49
31	7.96	9.94	5.1	0.64	75	8.02	9.46	5.9	0.16	119	7.88	8.59	5.7	0.71
32	7.89	9.61	5.2	0.59	76	8.04	9.52	5.5	0.18	120	7.95	9.78	5.9	0.85
33	7.96	9.68	5.4	0.58	77	8.01	9.77	5.6	0.18	121	7.95	9.69	5.7	0.20
34	7.87	9.79	5.4	0.66	78	8.06	9.41	6.0	0.19	122	7.88	9.23	5.5	0.15
35	7.87	9.78	5.7	0.65	79	8.03	9.90	5.9	0.19	123	7.93	9.11	5.5	0.19
36	7.94	9.78	5.5	0.59	80	7.96	9.77	5.5	0.23	124	7.89	9.78	5.5	0.85
37	7.84	9.78	4.7	0.59	81	8.00	9.72	5.3	0.20	125	7.88	9.23	5.7	0.69
38	7.90	9.76	4.8	0.61	82	7.92	9.71	5.6	0.21	126	7.84	9.03	4.9	0.28
39	7.86	9.79	4.6	0.66	83	7.88	9.33	4.5	0.24	127	7.91	9.10	5.1	0.22
40	7.91	9.81	3.4	0.64	84	7.89	9.66	5.3	0.20	128	7.98	9.17	5.2	0.16
41	7.89	9.82	4.1	0.61	85	7.90	9.23	5.3	0.23	129	7.93	9.33	5.5	0.13
42	7.87	9.83	4.7	0.58	86	7.90	8.86	4.4	0.23	130	7.90	8.42	5.7	0.43
43	7.84	9.81	5.2	0.60	87	7.87	8.83	4.1	0.21					
44	7.89	9.81	4.5	0.54	88	7.90	8.81	4.1	0.12					

The processed relationship figures from the three methods are shown in Figures 1, 2 and 3.

(1) The same starting point method:

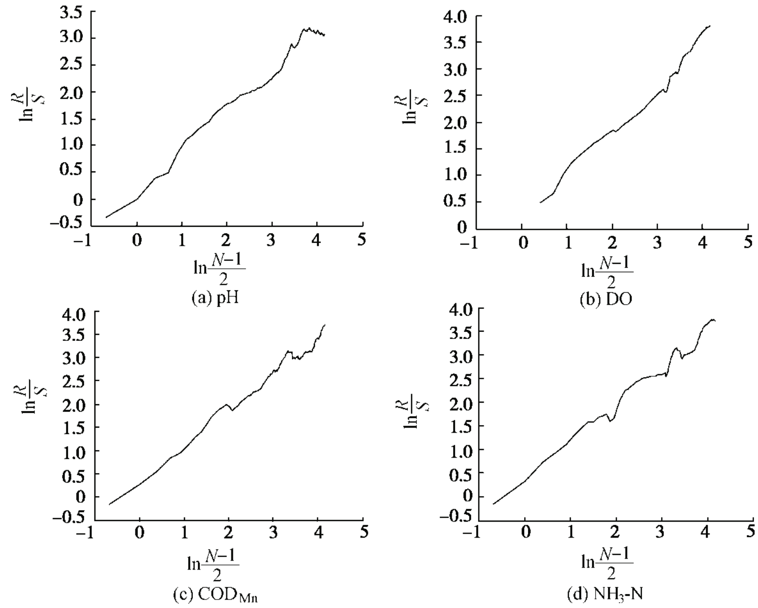

Figure 1 Hurst index obtained with the same starting point method

(2) The striding averaging method:

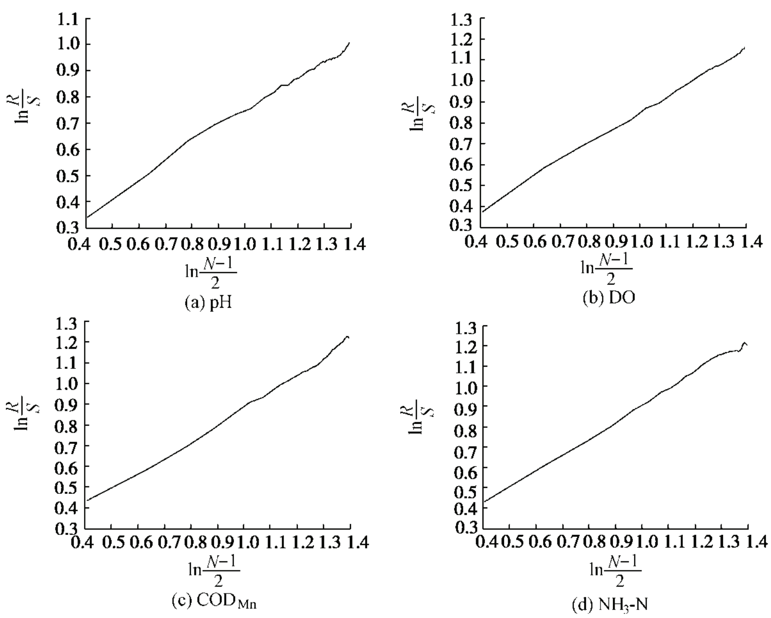

Figure 2 Hurst index obtained with the striding averaging method

(3) The stagger phase averaging method:

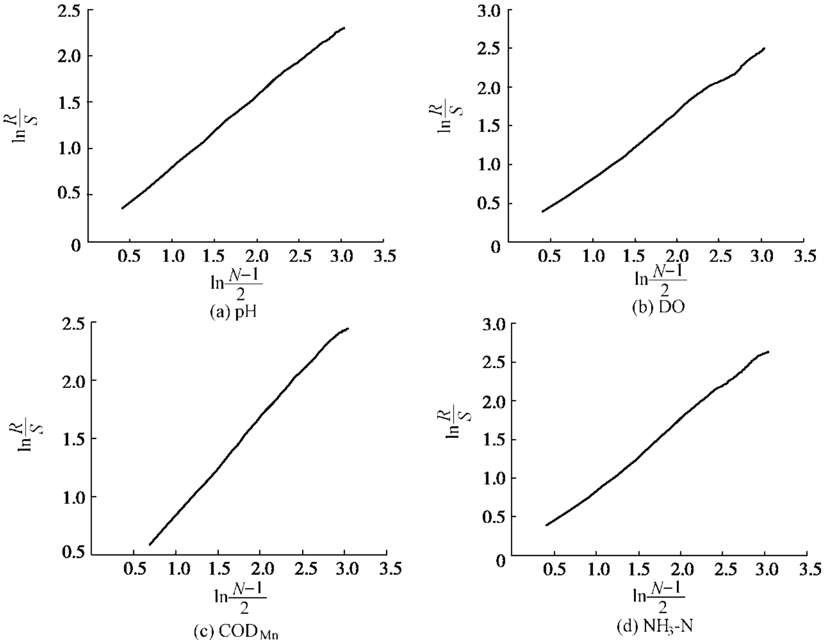

Figure 3 Hurst index obtained with the stagger phase averaging method

According to the regularities of the curves shown in these figures, and from the perspective of statistical regularity of all the R/S analytical methods, the stagger phase averaging method had the best regularity in this environmental assessment. This method is recommended for calculating the Hurst index of a pollutant time series. Here the Hurst index of the time series of water quality at Shazhu monitoring station on Taihu Lake in Wuxi, Jiangsu Province can be obtained by combining all the figures and using the least square method. The result is $H = 0.78$.

6 Conclusions

Based on this research on the fluctuation and randomness of the time series of each pollutant in the environmental quality assessment and the Hurst index of the Shazhu monitoring station on Taihu Lake in Wuxi, Jiangsu Province, and taking the three different computing methods into consideration, this study reached the following conclusions: as far as the stability of the Hurst index is concerned, the stagger phase averaging method gives the best result, the starting point method gives the worst, and the striding averaging method is in between. The Hurst index is an indicator of the stability of the pollutant concentration and of phenomena such as randomness, trend variation, and circulation sustention of pollutant concentration in the water. Therefore, the stagger phase averaging method is recommended for calculating the Hurst index of pollutant time series in order to make the evaluation of environmental quality more reasonable and more scientific.

References

Anh, V., Lunney, K., and Peiris, S. 1997. Stochastic models for characterization and prediction of time series with long-range dependence and fractality. *Advances in Architecture International Series*, 12(1), 67-73.

Castillo, O., and Melin, P. 2001. Hybrid intelligent systems for time series prediction using soft computing techniques and fractal theory. *International Journal of Smart Engineering System Design*, 3(3), 187-194.

Hambly, B. M., and Jones, O. D. 2003. Thick and thin points for random recursive fractals. *Advances in Applied Probability*, 35(1), 251-277.

Ikeda, Y., Tokinaga, S., and Miyazaki, A. 1999. Estimation of fractal surfaces using scale expansion and its applications. *Electronics and Communications in Japan, Part III, Fundamental Electronic Science*, 82(9), 10-17.

Jaw, S. Y., and Chen, C. J. 1999. Near-wall turbulence modeling using fractal dimensions. *Journal of Engineering Mechanics*, 125(7), 804-811.

Kouzani, A. Z., He, F., and Sammut, K. 1999. Face image matching using fractal dimension. *IEEE International Conference on Image Processing*, 63(4), 642-646.

Liebovitch, L. S. 1997. Introduction to fractals with examples of biomedical applications. *International Journal of Smart Engineering System Design*, 1(1), 15-31.

Ni, H. Y., and Liu, X. L. 2005. Fractal characteristics and R/S analysis of time series of natural hazards. *Journal of Natural Disasters*, 14(6), 37-41. (in Chinese)

Ostrosi, E., Ferney, M. and Garro, O. 2003. A fractal-based approach for concurrent engineering. *Concurrent Engineering*, 11(4), 249-265.

Shao, H., Shi, Z. R., and Zhao, Q. X. 2006. Fractal features analysis on correction dimension of accidents. *Systems Engineering-Theory & Practice*, (4), 141-144. (in Chinese)

Shen, X. H., and Chen, J. H. 2005. Application of fractal and chaos theory in turbulence study. *Journal of Henan University of Science and Technology (Natural Science)*, 26(1), 27-30. (in Chinese)

Wang, W. S., Xiang, H. L., and Zhao, D. 2005. Estimating the fractal dimension of hydrological time series by wavelet analysis. *Journal of Sichuan University (Engineering Science Edition)*, 37(1), 1-4. (in Chinese)

Weffimuny, R., and Pehumadu, D. 2003. Automated digital image based measurement of boundary fractal dimension for complex nanoparticles. *Particle and Particle Systems Characterization*, 20(1), 18-24.

Zhang, S. W., Wang, W. S., Ding, J., and Chang, F. X. 2005. Application of fractal theory to hydrology and water resources. *Advances in Water Science*, 16(1), 141-146. (in Chinese)