Fractal behavior in the records of water pollution in Poyang Lake Inlet

Lili Wang¹,²,³

¹School of Science, Nanchang Institute of Technology, Nanchang Jiangxi 330099, China
²Jiangxi Province Key Laboratory for Water Information Cooperative Sensing and Intelligent Processing, Nanchang Institute of Technology, Nanchang Jiangxi 330099, China
³Email: wlllzh2008@126.com

Abstract. Based on the weekly average pH, DO, COD and NH₃-N monitoring data in Poyang Lake Inlet during the period from 2016 to 2017, we investigated their temporal by using an integrative approach combining the phase space reconstruction and correlation dimension. The correlation dimension $d$ values showed that the water pollution dynamic is a complex and chaotic system, and its complexity decreases along with the increase of temporal scale. The number of effective degree of freedom (DOF) of dynamic system reflected by the weekly average pH, DO, COD and NH₃-N time series are 7, 5, 5 and 4 respectively. To describe the pH pollution dynamics, they need at least 2 to 5 independent variables respectively at weekly scale, whereas DO, COD and NH₃-N pollution need at least 2 to 3 independent variables.

1. Introduction

Water pollution is one of the most important environmental problems which have been a problem obsessed all over the world [1]. Many scholars from various countries have done a lot of research work on how to improve the validity and accuracy of water pollution prediction, and also proposed various forecasting methods. However, at present, it is still very difficult to achieve the desired forecast target. The main reason for this lies in that the factors affecting water pollution are not only the distribution and discharge of pollution sources, but also many complicated factors, such as hydrological conditions, migration and transformation of pollutants, complex terrain features, etc. [2]. What’s more, these factors are not independent of each other and they generate complex nonlinear interactions at various spatial and temporal scales. So the temporal evolution of water pollution exhibits complex nonlinear characteristics. If appropriate nonlinear mathematical physics methods would be applied to study some important dynamic characteristics of water pollutants’ temporal evolution, so as to understand its intrinsic principals and mechanisms.

The water pollution is a complex system with many pollutants [3]. Furthermore, there is a strong nonlinear interaction of the internal behavior of the system. And the ways of system feedback and adjustment are various and changing all the way, and the system components and its related processes also exhibit high spatial and temporal heterogeneity. From the complexity theory, the water pollution process is an open, dissipative systems and a complex phenomena, which has become more complicated and diverse due to the strengthening of human activities. Its formation and evolution show that it is a macroscopic and holistic behavior of a certain factor, and it is also a comprehensive,
multi-factor behavior. At present, the theoretical methods describing the complexity of dynamic systems mainly include self-organization theory, fractal geometry, chaos theory, renormalization group theory, etc [4,5]. Among them, the fractal analysis, as one of the nonlinear methods to deal with complex phenomena, has been applied in many research fields, such as earthquakes [6,7], climate change [8,9], air pollution [10,11], DNA strands [12], heart rate [13], stock market [14] and impervious surfaces [15]. However, water pollution dynamic has been identified as a set of water environment states of a dynamical, chaotic system showing deterministic variability [16]. The interpretation of climate as a complex intervariable organization is a key issue for understanding temporal evolution of dynamic systems [9,17]. Subsequently, valuable applications of chaos and nonlinear dynamical systems theory in water pollution science have been found [18,19]. The idea of pollution evolving on a strange attractor is often invoked, sometimes implicitly, to illustrate qualitative aspects of theories and conjectures regarding the water environment system’s behaviour [20]. Many scholars had detected the water pollution dynamic whether there exists a chaotic attractor and it is resulted from evolution of the non-linear chaotic dynamic system. And chaos and certainty of water pollutants time series, such as pH, DO, COD and NH3-N, from different time scales, such as 15 seconds, 15 minutes, an hour, a day, a week, two weeks or even a month. For example, Zhou et al (2011) has pointed out that there is nonlinear chaotic behavior in NH3-N pollution system [16].

In this paper, Poyang Lake Inlet (Hamashi environmental monitoring station) water pH, DO, COD and NH3-N series are chosen as research objects. Poyang Lake (115º47′~116º45′E, 28º22′~29º45′N) is the largest freshwater lake in China [21]. Poyang Lake Inlet water quality data are provided freely on the Internet by Ministry of Environmental Protection the People’s Republic of China web site: http://www.mep.gov.cn/. We have used weekly average pH, DO, COD and NH3-N monitoring data of Poyang Lake Inlet and Outlet from 2016 to 2017. The study area and monitoring sections are shown in Figure 1.

In order to understand the complexity of water pollution dynamics in Poyang Lake, China, this study investigated the pH, DO, COD and NH3-N dynamics from multiple temporal scale and spatial perspectives by using a comprehensive approach including the phase space reconstruction and correlation dimension.

![Figure 1. Location of the study area and monitoring sections.](image)

### 2. Method

For a given system, let its ordered output sequence with some state variable of evolution over time $\{Y_t\}$: $y(t_i), i=1,2,\ldots,N$, where $N$ is sample size or sequence length. The steps to solve correlation dimension by the phase space reconstruction was performed as described by Grassberger and Procaccia[22].
A sufficient long time data can be described as some trajectories in the phase space of the dynamic system, which can reflect an asymptotic behavior of a certain attractor in the dynamical system. The dimension $d$ of the attractor is usually smaller than the dimension $m$ of phase space, and $m$ is generally unknown. To characterize the attractor by the series, an appropriate phase space must be established, and then the attractor is embedded in it.

Since time series data are discrete temporal variables, Ruelle (1981) have proposed to use discretized time series and $(m-1)$ drift[23]. That is

$$\bar{X}_i = (x(t), x(t + \tau), \cdots, x(t + (m-1)\tau)) \quad i = 1, 2, 3, \cdots$$

It can replace $\{x_i\}$ under continuous variables, where $\{x_i\}$ is standardized $\{Y_i\}$, $\tau$ is time delay parameter. So the system attractor can be embedded in the phase space supported by a variable time series and its extended (or drifted) coordinates. It should be noted that the correlation integrals are defined to distinguish between stochastic and chaotic behaviors for the set of points on the attractor [24]. In order to characterize the structural features of chaotic attractors, the relative sizes of $\tau$ should guarantee that the components of the $m$-dimensional phase space are linearly independent. So, these points $\bar{X}_1, \bar{X}_2, \bar{X}_3, \cdots$ in the phase space form a chaotic orbit. In this paper, $\tau=1$.

Taking different upper limit $l$ values to calculate its corresponding correlation function $C(l,m)$

$$C(l,m) = \frac{1}{n^2} \sum_{i,j=1 \atop i \neq j} \theta(l - \|\bar{X}_i - \bar{X}_j\|)$$

Where $\theta$ is Heaviside function

$$\theta(r) = \begin{cases} 1, & r > 0 \\ 0, & r < 0 \end{cases}$$

It is worth mentioning that $C(l,m)$ is a cumulative distribution function, which describes the probability that the distance $\|\bar{X}_i - \bar{X}_j\|$ between two points in the phase space is less than $l$. If the $l$ value is too small, or the distance exceeds $l$, then $C(l,m)=0$; If the $l$ value is too large, in other words, the distance between any two points is less than $l$, then $C(l,m)=1$. So the correlation function $C(l,m)$ can reflect the nature of the attractor only if the $l$ is appropriate.

According to the definition of the associated dimension $C(l,m)\sim l^d$, the slope of the straight line on the curves is the correlation dimension $d$ of the attractor.

$$d = \frac{\ln C(l,m)}{\ln l}$$

Repeat steps (1)–(3) through $m\equiv m+1$. If these slopes tend to be stable as the embedded dimension $m$ increases, the dimension of the attractor can be obtained. Otherwise, there is no chaotic attractor in the system.

In this paper, under different embedded dimension $m$, the slopes of the 5-point least squares line are obtained point by point, and the average value of $d_\rho$ within 30% fluctuation range near the maximum $d_\rho$ is correlation dimension of a given embedded dimension $m$. Continuing upper the embedded dimension $m$, when $(d(m)-d(m-1))/d(m)\leq10\%$, $m$ is considered to be a saturated embedded dimension $m_s$, and $d(m)$ is the corresponding correlation dimension $d_s$.

What’s more, for a given observation data, when $n$, $m$ and $l$ is used as parameters for numerical calculation, there are three theoretical requirements as following. Firstly, $N$ or the length of the sequence is large enough to ensure that the phase space trajectories evolve on the attractor and cover it. Secondly, the embedded dimension $m$ is large enough to ensure that the attractor can be embedded and calculate the saturated fractal dimension $d_s$, which reflects the structural characteristics of the attractor and describes the number of minimum state space variables for the chaotic attractor. According to the Whitney’s theorem [25], the saturated embedded dimension and the saturated fractal dimension satisfy $m_s=2d_s+1-d_s^2$, but there is a constraint between $m_s$ and $n$. Thirdly, the upper limit $l$ should be small.
enough to ensure the accuracy of the computation. If \( l \) is too small or too large, this caused the \( d \) calculation to distort. So the appropriate range of \( l \) values is selected to get meaningful results.

3. Results and discussion

Based on fractal theory and the above-mentioned calculation method of correlation dimension, the fractal dimensions of Poyang Lake Inlet water pH, DO, COD and NH\(_3\)-N time series were calculated (Table 1~2 and Figure 2~5).

Table 1. Fractal dimension of average pollutants in Poyang Lake at weekly scale.

| pollution index | \( d \) \((m=2)\) | \( d \) \((m=3)\) | \( d \) \((m=4)\) | \( d \) \((m=5)\) | \( d \) \((m=6)\) | \( d \) \((m=7)\) | \( d \) \((m=8)\) | \( d \) \((m=9)\) | \( d \) \((m=10)\) | \( d \) \((m=11)\) | \( d \) \((m=12)\) | \( d_s \) | \( m_s \) |
|-----------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| pH              | 1.7221           | 2.3763           | 2.8560           | 3.2793           | 3.6735           | 3.9004           | 4.081            | 4.1723           | 4.2995           | 4.5111           | 4.7859           | 3.9004           | 7                |
| DO              | 1.3463           | 1.6225           | 1.8665           | 2.0584           | 2.1539           | 2.2119           | 2.2590           | 2.3042           | 2.3621           | 2.4177           | 2.4680           | 2.0584           | 5                |
| COD             | 1.3423           | 1.5667           | 1.8095           | 2.0207           | 2.1430           | 2.2298           | 2.2842           | 2.3402           | 2.3960           | 2.4680           | 2.4937           | 2.0584           | 5                |
| NH\(_3\)-N      | 1.4286           | 1.6739           | 1.8214           | 1.9364           | 1.9860           | 2.0374           | 2.0412           | 2.0412           | 2.0412           | 1.9814           | 1.9554           | 1.9225           | 1.8213           |

Table 2. The statistics for fractal dimension of average pollutants in Poyang Lake at weekly scale.

| pollution index | average value | median | standard deviation | minimum value | maximum value | range | kurtosis | skewness |
|-----------------|---------------|--------|--------------------|---------------|---------------|-------|----------|----------|
| pH              | 3.6298        | 3.9004 | 0.9116             | 1.7221        | 4.7859        | 3.0638 | 0.2612   | -0.9502  |
| DO              | 2.0941        | 2.1829 | 0.3359             | 1.3463        | 2.4680        | 1.1217 | 0.9599   | -1.1912  |
| COD             | 2.01822       | 2.03475| 0.3367             | 1.3423        | 2.4893        | 1.1470 | 0.0635   | -0.6053  |
| NH\(_3\)-N      | 1.8848        | 1.9459 | 0.1792             | 1.4286        | 2.0412        | 0.6126 | 3.1390   | -1.7607  |

From Tables 1, 2 and Figure 2, the correlation dimension of pH time series tends to be stable when embedded dimension \( m \geq 7 \), and its correlation dimension \( d \) is 1.7221~4.7859 corresponding to its saturated embedded dimension. So the number of effective DOF of dynamic system reflected by the weekly average pH time series is 7. If its changing characteristics need to be properly describe, dynamic system modeling requires at least 2 to 5 independent state variables.

From Tables 1, 2 and Figure 3, the correlation dimension of DO time series tends to be stable when embedded dimension \( m \geq 5 \), and its correlation dimension \( d \) is 1.3463~2.468 corresponding to its saturated embedded dimension. So the number of effective DOF of dynamic system reflected by the
weekly average DO time series is 5. If its changing characteristics need to be properly describe, 
dynamic system modeling requires at least 2 to 3 independent state variables at weekly scale.

![Graph of lnC(r) versus ln(r) and correlation dimension (d) versus embedded dimension (m)](image)

**Figure 3.** A plot of lnC(r) versus ln(r) (right) and the correlation dimension (d) versus embedded dimension (m) (left) for DO time series.

From Tables 1, 2 and Figure 4, the correlation dimension of COD time series tends to be stable 
when embedded dimension \( m \geq 5 \), and its correlation dimension \( d \) is 1.3423—2.4893 corresponding to 
its saturated embedded dimension. So the number of effective DOF of dynamic system reflected by 
the weekly average COD time series is 5. If its changing characteristics need to be properly describe, 
dynamic system modeling requires at least 2 to 3 independent state variables at weekly scale.

![Graph of lnC(r) versus ln(r) and correlation dimension (d) versus embedded dimension (m)](image)

**Figure 4.** A plot of lnC(r) versus ln(r) (right) and the correlation dimension (d) versus embedded dimension (m) (left) for COD time series.

From Tables 1, 2 and Figure 5, the correlation dimension of NH3-N time series tends to be stable 
when embedded dimension \( m \geq 4 \), and its correlation dimension \( d \) is 1.4286—2.0412 corresponding to 
its saturated embedded dimension. So the number of effective DOF of dynamic system reflected by 
the weekly average NH3-N time series is 4. If its changing characteristics need to be properly describe, 
dynamic system modeling requires at least 2 to 3 independent state variables at weekly scale.
Figure 5. A plot of $\ln C(r)$ versus $\ln(r)$ (right) and the correlation dimension ($d$) versus embedded dimension ($m$) (left) for NH$_3$-N time series.

4. Conclusions
Summarizing the above results, we elicited the conclusions as follows:

Different embedded space dimensions of weekly average pH, DO, COD and NH$_3$-N time series all have corresponding scalefree intervals, and their slopes of the straight lines portion of correlation function all tend to stabilize as the embedded dimension $m$ increases. That proves that water pollution system is a chaotic attractor.

Water pollution system contains multiple levels, and high levels (small dimension values) is a low level’s (large fractal dimension values) background, that is, large pollution control small pollution. The change rule in water pollution is more easily found at the high levels. There are many uncertain factors at a low levels, which its effection on high levels can be understood as a damping and it is a feedback on the high level of control. So we can probably say that a small water pollution events change large water pollution events.

The number of effective DOF of dynamic system reflected by the weekly average pH, DO, COD and NH$_3$-N time series are 7, 5, 5 and 4 respectively. If the pH, DO, COD and NH$_3$-N pollution dynamics are properly describe, pH dynamic need at least 2 to 5 independent state variables for pH at weekly scale, while DO, COD and NH$_3$-N time series need at least 2 to 3.

Acknowledgments
This work was supported by the Science and Technology Research Project of Jiangxi Provincial Education Department (GJJ180926), the Open Research Fund of Jiangxi Province Key Laboratory of Water Information Cooperative Sensing and Intelligent Processing (2016WICSIP014).

References
[1] Li T and Wu X F 2010 Environ. Sci. 31 2619-2626
[2] Minella M, Maurino V, Minero C and Vione D 2013 Int. J. Environ. An. Ch. 93 1698-1717
[3] Benedetti L, Blumensaat F, Bönisch G, Dirckx G, Jardin N, Krebs P and Vanrolleghem PA 2005 Water Sci. Technol. 52 171-179
[4] Malderbrot BB 1967 Science 156 636-638
[5] Ostwald M J and Wassell S R 2002 Nexus Netw. J. 4 123-129.
[6] Hayakawa M, Ito T and Smirnova N 1999 Geophys.Res.Lett. 26 2797-2800
[7] Liu Z H, Wang L L and Qi S H 2018 J. Phys.: Conf. Ser. 1053 012061
[8] Yédjinnavènè Ahokpossi 2019 Hydrolog. Sci. J. 63 1-27
[9] Liu Z H, Xu J H, Chen Z S, Nie Q and Wei C M 2014 Stoch. Env Res. Risk A. 28 1383-1400
[10] Liu Z H, Sun L N, Wang J Y, Wang L L, Shi K and Liu C Q 2017 Fresen. Environ. Bull. 26 7681-7686
[11] Liu Z H, Wang L L and Zhu H S 2015 Atmos. Pollut. Res. 6 457-486.
[12] Nicolay S, Brodie EBBO, Touchon M and Audit B 2007 Phys. Rev. E 75 032902.
[13] Perakakis P and Taylor MNE 2009 Biol. Psychol. 82 82-88
[14] Samadder S, Ghosh K and Basu T 2013 Fractals 21 1350003
[15] Nie Q, Xu J H and Liu Z H 2014 Earth Sci. Inform. 8 1-12
[16] Zhou W, Zhang B, Liu Z H, Deng Q C, Qin F, Di B F and Yu B 2011 Bull. Soil Water Conserv. 31 138-141
[17] Liu Z H, Xu J H and Shi K 2014 Theor. Appl. Climatol. 115 685-691
[18] Beck M B 1987 Water Resour. Res. 23 1393-1442
[19] Wu J, Lu J and Wang J Q 2009 Env. Model. Software 24 632-636
[20] Holmes P and Marsden J 1981 Arch. Rational Mech. Anal. 76 135-166
[21] Li Y T, Deng J Y and Sun Z H 2003 Int. J. Sediment Res. 18 138-147
[22] Grassberger P and Procaccia I 1983 Phys. Rev. Lett. 50 346-349
[23] Ruelle D 1981 Commun. Math. Phys. 82 137-151
[24] Xu J H, Chen Y N, Li W H, Liu Z H, Tang J and Wei C M 2016 Theor. Appl. Climatol. 123 1-13
[25] Whitney H 1932 Congruent graphs and the connectivity of graphs Amer. J. Math. 54 160-168