Parameter selection for phase space reconstruction in hydrological series and rationality analysis of its chaotic characteristics

Hongyan Li, Shanshan Bao and Yunqing Xuan
Abstract. This study performed a rationality analysis of the delay time and embedding dimension value during phase space reconstruction in hydrological series and the effect on their chaotic characteristics. Using a monthly average runoff time series from the Ayanqi an station (upstream) and the Jiangqiao station (midstream) in the Nen River Basin, we reached the following regularity conclusions. ① Based on the flood season (4 months) in the Nen River Basin, we can deduce that the correlation sequence length for the runoff is 4~5 months, i.e., the delay time =3 or 4 is a reasonable choice. ② Learn from the predictability experiment results for the monthly rainfall time series, we know that the calculation results of the G-P algorithm for the dimension of runoff series for the Nen River Basin are reasonable, i.e., the embedding dimension is no more than seven. ③ the most suitable parameters for the phase space reconstruction and its chaotic characteristic index in the Nen River Basin are as follows: delay time = 3–4, embedding dimension = 6–7, correlation dimension = 2.90–3.00, maximum Lyapunov index = 0.24–0.32, and the forecast time is 3–4 months.

Keywords: Chaos Theory ; Delay Time ; Embedding Dimension ; Phase Space Reconstruction

1. Introduction

The classical theory of hydrology considers that hydrological phenomena have deterministic and random properties: as the astronomical and macroscopic geographical factors are stable, the hydrological phenomena have yearly periodicity and obvious seasonality; while hydrological phenomena are affected by many secondary and indirect factors, such as changes in atmospheric circulation, the temporal and spatial distribution of precipitation, and the evapotranspiration process, the mechanisms of which are varied and complex. However, no factor plays the leading role, which eventually leads to the randomness of hydrological phenomenon [1]. This randomness is essentially external, i.e., deterministic systems introduce random initial conditions, random parameters, random external forces (noise), etc., which makes the system output is no longer deterministic [2]. In 1963, however, the American meteorologist Lorenz detected the extreme instability of numerical equations to the initial value in a numerical experiment, i.e., the intrinsic randomness of deterministic equations [3]. In order to explain this “chaotic phenomenon” known as “The Butterfly Effect,” scientists have formulated the mathematical results as Chaos Theory.

The application of Chaos Theory in hydrology can be divided roughly into two categories, i.e.,
estimating the predictability scale of hydrological series [4-6] and combining a variety of mathematical models, such as similar point methods [7], regression analysis [8], wavelet analysis [8], fuzzy mathematics [9], neural nets [10-12], radial basis function [13], and support vector machines [14] to construct medium- and long-term hydrologic forecasting model. In addition, studies have also used the chaotic characteristics of hydrological series to reduce the hydrologic data noise [15] and perform hydrologic data interpolation [16]. These studies are all based on the reconstruction of phase space in hydrological series and the analysis of their chaotic characteristics. The application and existing problems in Chaos Theory have been systematically and fully discussed previously [17], so they won't be covered again here.

This article discusses the determination of the delay time and embedding dimension when applying Chaos Theory to the reconstruction of phase space for a hydrological series and its chaotic characteristics, where we highlight the physical cause analysis of hydrology. We reconstruct the phase space $Y$ after determining the delay time $\tau$ and the embedding dimension $m$ of nonlinear system $X$ and identify $Y$ based on its chaotic characteristics, i.e., we calculate the maximum Lyapunov index and judge whether this system has chaotic characteristics, before further analysis of its predictability and predictable time scale.

2. Phase space reconstruction for a time series

In mathematics and physics, the phase space is a representation system used to describe the system status and evolution process, where any status of the system has a corresponding phase space point. If a dynamic system is expressed as a group of ordinary differential equations, we can establish a clear link between the space coordinates and the state variables, before deducing the property of attractors directly to evaluate the chaotic characteristics of the dynamic system. However, if the movement differential equation of a reality system is unknown, it is very difficult to get the attractor characteristics of the system, so we have to reconstruct a phase space by observing the sequence $x(t)(t = 1,2,\cdots,N)$. The phase space reconstruction theory of Packard [18] introduces Chaos Theory into time series analysis, which considers that the time series themselves include the relevant information for all of the dynamic system variables. By analyzing the observational data, we can change the observations of some fixed time delay points into new coordinates that jointly determine a point in multi-dimensional state space, so the phase space of the system can be reconstructed from a variable time series. Takens’ embedding theorem [19] developed the Packard’ thoughts and proposed a lower bound for the embedding dimension in phase space as

$$m \geq 2D + 1$$

where $D$ is the strange attractor dimension of the chaotic system, known as the correlation dimension. The $m$ -dimensional vector sequence is reconstructed using $x(t)(t = 1,2,\cdots,N)$

$$X(t) = \{x(t), x(t + \tau), \cdots, x(t + (m-1)\tau)\}^T \quad (t = 1,2,\cdots,M) \quad (1)$$

where $M = N - (m-1)\tau$ and $\tau$ is the delay time. The $m$ -dimensional state space that is produced by the observed value $x(t)$ and its delay value $\tau$ is the phase space reconstruction. Based on Takens’ embedding theorem, we know that provided we choose a suitable “delay time $\tau$ ” and “embedding dimension $m$”, the trajectories of the phase space reconstruction in the embedding space and the original system are dynamically equivalent in a diffeomorphic sense. The phase space reconstruction maintains the original topology structure and it possesses the same dynamic characteristics. And then we can analyze the chaotic characteristics of the original dynamic system in the phase space reconstruction.

3. The effect of the delay time on the phase space reconstruction and its chaotic characteristics

For the high sensitivity of phase space reconstruction to the delay time $\tau$, data experiments have been conducted in the Lorenz system (Lorenz Equation) and the Luo Sile system (Rössler
Equation), where the results showed that \(20-22\): if \(\tau\) is too small, the constructed attractor is extruded into the vicinity of the diagonal line of the coordinates system; whereas if \(\tau\) is too large, the trajectory of data points wrinkles and folds, so it is difficult to obtain clear projection relationships. The choice of the delay time is essential for the reconstruction of phase space and its chaotic characteristics.

### 3.1 General method for deriving the delay time and the existence problem

The influential conclusions about delay time selection for the phase space reconstruction in the above text are obtained using differential equations where their sample size is infinite and the sample lacks noise, which should be treated with extreme caution during the direct popularization of actual systems. During the chaotic analysis of hydrological series, the derivation of the delay time \(\tau\) has been discussed systematically in the literature\(^{23}\) but confusion still exists, i.e., first, the samples are limited in the runoff sequence; second, the samples are rich in noise (the permissible relative testing error for flow is 5\%) \(^{24}\). In hydrology, we also need to determine whether the correlation coefficient \(r\) is greater than 0.8 to judge the correlation of the two indicators, instead of 0 in the mathematical sense. Previous studies \(^{25-26}\) have used \(r = 0.1\) and 0.5 as the criterion for deducing the delay time \(\tau\) depending on the characteristics of problems in their respective areas. Therefore, when deducing the delay time, we should start from the physical significance of the delay time required for the phase space reconstruction, i.e., “keeping independent or reaching minimum correlation among the embedding coordinates” \(^{10}\) where the “minimum correlation” should be based on the actual problem. Jumping from the simple mathematical concept and applying the physical interpretation of chaos theory to a real problem are the ways of exploring the deduction of the delay time.

### 3.2 The effect of the delay time on the phase space reconstruction and its chaotic characteristics

We adopt the commonly used experimental methods of Chaos Theory \(^{27}\) to study the effect of different delay time to a hydrological time series and its chaos characteristics. It is difficult to estimate a suitable delay time \(\tau\) and the embedding dimension \(m\) for a sequence, but we may list the possible values of \(\tau\), before deducing the corresponding \(m\) and the maximum Lyapunov index \(\lambda_1\). We can use the methods for deriving \(m\) and \(\lambda_1\) that have been proposed in the literature \(^{27-28}\). We used the monthly averages of the runoff sequences in the upper reaches of Ayanqian station and the midstream Jiangqiao station in the Nen River Basin during 1952 to 2005, as well as the monthly averages in the upper reaches of Tangnaihe station and the midstream Longmen station in the Yellow River during 1953 to 2003, and selected the delay time \(\tau = 1-24\) to calculate the phase space reconstruction and the chaotic characteristic index. When drawing \(m\) and \(\lambda_1\) in the same graph, a normalization process should be applied to eliminate the effect of dimension, i.e., the respective volume indicators divided by their respective maximum value, which is known as the “coefficient of chaos index model ratio.” The results are compared in Figures 1 and 2.
Figure 1: The effect of the delay time on the Nen River runoff sequence during the phase space reconstruction and its chaotic characteristics.

For the Nen River Basin, with the change of $\tau$, $m$ has a sudden change at the points of $\tau = 12 (m = 40)$ and $24 (m = 20)$, while at any other time it fluctuates lightly. This situation for the Yellow River runoff is not so obvious, although its calculated value reaches the maximum at the points of $\tau = 12 (m = 40)$ and $24 (m = 20)$ with slightly change. For the two basins, the change of $\tau$ has little effect on the upstream sections (Ayanqian and Tangnaihe), whereas it is large in the downstream sections (Jiangqiao and Longmen).

For the Nen River Basin, $\lambda_1$ exists obvious periodicity with $\tau$ has a peak at integer multiples of three and a valley at integer multiples of four. The value of $\lambda_1$ is significantly higher in the upstream section than the downstream section. For the Yellow River Basin, $\lambda_1$ exists obvious periodicity in the upstream of Tang Naihe station. It is slightly different from the Nen River Basin, although it has a peak at integer multiples of three, when $\tau = 6, 12, \text{and } 18$, the value of $\lambda_1$ is much higher than the value at $\tau = 9, 15, \text{and } 21$. The change of $\lambda_1$ in the downstream of Longmen station does not exhibit significant regularity and the changing trends are also inconsistent in the upstream and downstream.

Based on an intuitive overview of the two watersheds, a further analysis combined with the expert knowledge of hydrology is as follows.

1. From a randomness perspective, there are significant differences between the upstream and
downstream regions, where the former is less affected by human activity than the latter. There are significant differences between the Yellow River Basin and Nen River Basin, where the former is more affected by human activity than the latter. The construction of large water projects in the upper reaches of the Yellow River began in the 1950s, such as Liujiaxia, Yanguoxia and Qingtongxia which were started in 1958. By contrast, the Tang Naihe station, which is located in the river head, is less affected by human activity. The upstream of Nen River Basin has few water projects, except for the Nierji hydraulic key project, which began to store water in 2005. Because the data used in this article predate 2005, it can be considered the natural runoff condition. Thus, it is clear that runoff series with less influence of human activity have obvious chaotic characteristics. Whereas, it is unfit for chaotic analysis.

2. The value of $\tau$ (above) has a sudden change at 12 and 24, which corresponds to the explicit hydrological cycle (12 months). Based on this, we can deduce that the value of $\tau$ should be less than 12. In order to determine a reasonable value of $\tau$, we need to perform a further analysis of the various indicators of $\tau$ in the range of 1~12 segments. The main recharge of the Nen River Basin is precipitation, so the inflow is concentrated in the flood season, i.e., June 15 to September 15. The main recharge is precipitation during the flood season, whereas groundwater recharge is more important after the flood season. The autocorrelation of the runoff process does not have a causal relationship essentially, while the rainfall recharge capacity and the groundwater so have such a correlation with the runoff. Therefore, flood runoff and non-flood runoff certainly do not have a close correlation, which manifests as two correlated processes. In the Nen River Basin, the flood season lasts for 4 months so we can deduce that the correlation sequence of runoff is 4~5 months, i.e., $\tau = 3$ or 4 is a reasonable choice. The flood season in the above area of the Tangnaihe station on the Yellow River is from July to October. The runoff changes are mainly affected by rainfall and temperature, i.e., the rainfall affects runoff production directly while the temperature may affect the snowmelt runoff, and there is almost no influence of human activity. There is little precipitation from November to April in the next year, so the runoff is mainly composed of basin water supplies and melting snow. Generally, the runoff process presents a stable extinction rule from late October to November [29]. Therefore, the Yellow River Basin runoff has a more complex composition and variation. The middle and lower reaches of the Yellow River are seriously affected by human activities, so the runoff randomness of the natural environment is overshadowed by human factors. In natural conditions, the natural factors are relatively stable over long (or very long) time scales whereas the stability of human factors is relatively poor, so the operating program of the key control project may need to be adjusted quickly depending on the production requirements.

3. The effect of $\tau$ on $m$ is not significant, whereas there is a significant effect on the chaotic characteristics of the phase space constructed using different $\tau$ and $m$ values, i.e., the chaotic characteristics of phase space constructed using different combinations of $\tau$ and $m$ have significantly differences.

4. Effect of the embedding dimension on the phase space reconstruction and its chaotic characteristics

In order to analyze the effect of $m$ on the phase space reconstruction and its chaotic characteristics, Nen River Basin was selected as the study area, which has slightly human activity and simple watershed composition. The chaotic characteristic indexes were calculated by using $\tau = 3$ and 4 which are corresponding to different $m$ values.
Figure 3 (a) The effect of different embedding dimensions on the phase space reconstruction and its chaotic characteristics. (Ayanqian)

Figure 3 (b) The effect of different embedding dimensions on the phase space reconstruction and its chaotic characteristics. (Jianqiao)

Figure 3 shows that with the increasing of $m$, the trend of the curve $\lambda_i$ for the two stations is consistent at $\tau = 3$ and 4, where they all converge to the vicinity of 0.05. The identifiers “•” and “▲” indicate the embedding dimensions which are calculated by G-P algorithm and their position on the curve is not special. Take’s embedding theorem showed that in order to reconstruct the single variable time series into a multidimensional phase space, $m$ should be large enough.

$$m \geq 2D + 1$$

The dimension of strange attractors reflects the complexity of the attractor structure, while it also reflects the amount of information in the attractor, so it can also be used to characterize the
geometric structure of attractors. Correlation dimension $D$ is a common definition of the dimension, where the physical concept is clear and it has a strict mathematical definition. As a purely mathematical definition, however, there may be many difficulties if it is used to analyze the experimental data\cite{30}.

With respect to the experimental method of the value of $m$, based on the phase space modeling theory, literature \cite{31} performed a prediction experiment that is using the monthly rainfall time series and studied the forecast precision of the monthly rainfall along with the change of the embedding dimension $m$. It found that the phase space gave the best prediction results when $m=7$, so the maximum number of independent variables in the description of the chaotic system was no more than seven.

The runoff changes mainly depend on the combined influence of astronomy, meteorology, ocean, and the underlying surface. The conditions of the underlying surface are relatively stable for a certain basin, which may be considered to be a time-invariant system \cite{32}. Thus, the runoff impact factors we should consider are the basin precipitation and atmospheric circulation features. Based on the above predictability experiments related to the monthly rainfall time series, the calculation results of the G-P algorithm for the dimension $m$ are reasonable when using the monthly runoff time series of the Nen River Basin, i.e., Ayanqian station: $\tau=3$, $m=6$; $\tau=4$, $m=6$; Jiangqiao station: $\tau=3$, $m=7$; $\tau=4$, $m=6$.

5. The forecasting time scale for the runoff process

Predictability refers to the forecasting accuracy of the system in the near future or the long-term future and the feasibility of long-term forecasting is a highly controversial subject. Newtonian mechanics determinism considers that long-term forecasting is not only feasible but also does not limit the time, whereas others who hold the “Butterfly Effect” view believe that long-term prediction is impossible. However, Lin \cite{33} suggested that the chaotic behavior of chaotic systems is not a fully parametric domain and that the dynamic behavior of nonlinear chaotic systems is non-chaotic in most of the parameter domain, and that chaos can be used to make numerical predictions. The dynamic behavior of chaotic systems is rarely chaotic in part of the parameter domain and the chaos could not be used to make numerical predictions. Based on a large amount of model testing, the author reached the following conclusions.

(1) Theoretically, statistical methods are more rigorous than dynamic methods for the forecast of chaotic dynamic systems, whereas the opposite is true for the forecast of non-chaotic dynamic systems.

(2) It is difficult to determine the chaotic region parameters accurately for an objective system or to estimate whether a chaotic region exists, so it is more convenient and rigorous for statistical forecasting methods than dynamic systems.

(3) During numerical prediction, the methods used to make the average monthly forecast based on the average monthly data are more rigorous than when calculating the average monthly forecast based on instantaneous values or after calculating the average instantaneous value.

A previous study \[Lin, 1999a\] suggested that nonlinear dynamical systems are extremely sensitive to the initial conditions and the evolution of error manifests as indexed divergence, so the evolution behavior can be forecast objectively for a non-chaotic system or a chaotic system during a non-chaotic region. The reciprocal of the maximum Lyapunov index is used to define a system’s maximum possible forecasting time scale $T_f$, in previous studies, as follows:

$$T_f = \frac{1}{\lambda_1} \quad (2)$$

where $\lambda_1$ is the largest Lyapunov index.

The initial value sensitivity of chaotic system is such that if two trajectories are very close at first, they can diverge at an exponential rate in the phase space. The Lyapunov index is based on whether the phase trajectories have diffused movement characteristics to distinguish the chaotic
According to the calculational results in the above text, the most suitable parameters for phase space reconstruction and its chaotic characteristic index of the Nen River Basin should be as follows: delay time $\tau = 3 ~ 4$, embedding dimension $m = 6 ~ 7$, and the maximum Lyapunov index $\lambda_1 = 0.24 ~ 0.32$. These values can be used to estimate the forecast time $T_f$ for the Nen River Basin as 3~4 months.

6. Conclusions and outlook

This study proposed a general method from Chaos theory to study time series problems. We used the basin theorem for phase space reconstruction and its explanation in hydrology as the entry points, before we focused on the effects of the embedding dimension and delay time on the phase space reconstruction and its chaotic characteristics. We reached the following conclusions based on a monthly runoff time series for the control section of the upper and middle reaches of the Nen River Basin and the Yellow River Basin. From a randomness perspective, there are significant differences between the upstream and downstream sections, where the former is less affected by human's activity than the latter. It is clear that the chaotic properties of runoff processes are affected slightly by human activity are obvious. The flood season lasts 4 months in the Nen River Basin and based on this we can deduce that correlation sequence length for the runoff is 4~5 months, i.e., $\tau = 3$ or 4 is a reasonable choice. Based on the predictability experiment results for the monthly rainfall time series, the calculation results of the G-P algorithm for the dimension $m$ of runoff sequence in the Nen River Basin are reasonable, i.e., the embedding dimension is no more than seven. The most suitable parameters for the phase space reconstruction and its chaotic characteristic index in the Nen River Basin are as follows: delay time $\tau = 3 ~ 4$, embedding dimension $m = 6 ~ 7$, correlation dimension $D = 2.90 ~ 3.00$, maximum Lyapunov index $\lambda_1 = 0.24 ~ 0.32$, and the forecast time $T_f$ is 3~4 months.

Acknowledgments

We are grateful to the National Natural Science Foundation of China (NO:51379088), NSFC-NRF Joint Project “Study on regional water resources response and sustainable utilization under the changing environment” (51711540299), "13th Five-Year" Key Research Program “River lakes and reservoirs comprehensive control technology and application in SongLiao River basin" (2017YFC0406005) to provide financial support for this research.
References