Searching of Chaotic Elements in Hydrology

Sorin VLAD, Ştefan-Gheorghe PENTIUC
Stefan cel Mare University of Suceava, Romania
sorinv@seap.usv.ro, pentiuc@usv.ro

Abstract - Chaos theory offers new means of understanding and prediction of phenomena otherwise considered random and unpredictable. The signatures of chaos can be isolated by performing nonlinear analysis of the time series available. The paper presents the results obtained by conducting a nonlinear analysis of the time series of daily Siret river flow (located in the North-Eastern part of Romania). The time series analysis is recorded starting with January 1999 to July 2009. The attractor is embedded in the reconstructed phase space then the chaotic dynamics is revealed computing the chaotic invariants - correlation dimension and the maximum Lyapunov Exponent.

Keywords: big data, decision tree, Hadoop, MapReduce, pattern recognition.

I. INTRODUCTION

Nonlinear systems can exhibit a variety of behaviors including chaotic behavior. A nonlinear system is frequently very complex. Complexity determines a tight and subtle linkage among causes and effects, being so many causes and the global image created by the multitude of these causes (insufficiently known) being so large that the strict determination cause-effect is extremely difficult to made and almost impossible.

It was shown [1] that even simple deterministic systems can be influenced by the variation of some variables to produce complex (and possibly chaotic) behavior.

Chaos is a phenomenon that can occur only in nonlinear systems. The majority of natural systems (including here even the human being [2]) are nonlinear so we can draw the conclusion that chaos can arise almost everywhere in natural systems. The main benefits of proving the existence of chaotic dynamics in the behavior of a process is that short time forecasting is possible (with some degree of accuracy) and the long term forecasting is a pointless endeavor because one will obtain useless erroneous results. Beside that, chaos theory may offer new means of understanding the intrinsic nature of the system analyzed.

The main feature of a chaotic system its sensitivity to the variation of initial conditions. An infinitesimally small change of the starting values will influence dramatically the evolution in time of the system. The trajectories will diverge exponentially (the final values will be largely different). The divergence of nearby trajectories is measured by Lyapunov exponent (LE). A positive LE is a hint that the dynamics of the system is chaotic but it is not a sufficient condition. A positive LE alone isn’t proving that the system is chaotic since the measurement noise can determine LE to be positive. A negative LE indicates how rapidly the system restores its initial state after a perturbation.

Although it may seem a contradiction in terms, a chaotic system is ordered. The dynamics of such a system is deterministic (hence the order) meaning that if the initial values are repeated, the trajectories repeat identically as the first time. The order that governs a chaotic behavior can be “unveiled” by a proper representation in the state space. The state space dynamics can be visualized using return maps or phase portraits and should reveal the existence of a chaotic attractor. An attractor occurs due to the divergence of the trajectories in the phase space having a non integer dimension and exhibits self-similarity. The trajectories are dense, occupying a bounded region of the phase space and never repeating. An attractor will appear as a result of the stretching and folding of the phase space.

The possibility of predicting the evolution of a chaotic system is based on the following paradoxical situation: if a system is governed by deterministic laws and a complete description of the current state is available, then the future states are theoretically completely predictable. In practice is obviously that the current state will never be described with an infinite degree of precision and it’s also obvious that, as the nonlinearity increases, the output of the system is apparently erratic and long term unpredictable.

II. BACKGROUND AND RELATED WORK

Generally, the papers related to chaos theory applied in hydrology are dealing with one of the following classes of problems: investigation (and proving or not) the existence of chaos in river flows and water levels, building models for prediction and comparing the results obtained using chaos theory methods with the results obtained through other techniques (nonlinear modeling using Artificial Neural Networks and stochastic methods) [3][4].

In the past few years there is a tremendous interest to find ways of explaining and modeling the dynamics of the hydrological processes. It was not uncommon until recently to consider that these processes are influenced by a large set of variables. Stochastic hydrological processes modeling was also a common practice, implying the
opinion that the process is described almost completely by
the statistical features of a small set of parameters.[5]. The
model fitted is improved as new data become available [5].

The first paper investigating the existence of a strange
attractor in the time series of monthly rainfall was
published in 1987[6]. Researches has been carried out on
the effect of noise in hydrological time series [7] and on
the reliability of some chaotic invariants in this particular
case [8]. The existence of deterministic chaos was
investigated in the daily flow of one of the tributaries of
the river Po in Italy[9]. The dynamic of the variation of
high level of water at Venice lagoon was studied and
predicted in [10].

III. CHAOS IDENTIFICATION IN TIME SERIES

We will search for chaos signature in Siret river flow
time series (Figure 1) by following the steps required by
nonlinear time series analysis procedure. The subsequent
analysis will take into account a stationary time series (no
changes in mean and variance over time).

Figure 1. The preprocessed Siret flow time series.

A system is chaotic if there are fulfilled several
conditions: the autocorrelation function decays to 0, the
largest LE is positive and the correlation dimension
saturates as the embedding dimension increases.

Noise is an important factor which can influence
dramatically the analysis of a time series, affecting the
structure of an attractor. In order to avoid the contamination of data, a denoizing procedure must be
applied first. We used TISEAN software package to
perform this task.

The first step when trying to apply nonlinear analysis
techniques is the so-called reconstruction of the phase
space. According to Taken’s theorem, the missing
information about the dynamics of the system can be
obtained by embedding the available time series and
obtaining a set of delay coordinate vectors.

\[ X(n) = [x(n-(d-1))\tau, x(n-(d-2))\tau, ..., x(n-k), x(n)]^T \]  

1 where \( \tau \) is the time delay.

The reconstructed attractor formed by the lagged vectors
it’s a different morphism of the original attractor if \( m>2d \),
where \( d \) is the correlation dimension, and \( m \) is the
embedding dimension.

To choose the time delay, we applied mutual
information method which determine \( \tau \) as the first
minimum of mutual information function. It’s important to
determine the correct time delay because if it’s too small
no new relevant information is considered and if it’s too
large, the useful information is lost. Mutual information
between the realizations of a time series is given by:

\[ I(\tau) = \sum P(x(t_i), x(t_i + \tau)) \log \left( \frac{P(x(t_i), x(t_i + \tau))}{P(x(t_i)) P(x(t_i + \tau))} \right) \]  

2 where \( P(x(t_i)) \) is the probability density at \( x(t_i) \) and
\( P(x(t_i), x(t_i + \tau)) \) is the joint probability at \( x(t_i) \) and \( x(t_i + \tau) \).

The method used for computing the proper embedding
dimension is the method of false nearest neighbors (FNN)
which indicates that the number of FNN is dropping below
0.1% when the correct embedding dimension is attained
ensuring that the attractor is completely unfolded.

The correlation dimension is a metric computed with
Grassberg and Procaccia algorithm [11] and quantifies the
fractal geometry of the attractor by measuring the \( m \)
dimensional space filled by it, but also indicates the
minimum number of independent variables needed to
model the time series.

The correlation integral is defined as:

\[ C_m(\varepsilon) = \frac{1}{(T-m+1)(T-m)} \sum_{i<j} I_\varepsilon(x_i^m, x_j^m) \]  

3 Where \( I_\varepsilon = 1 \) if \( \|x_i^m - x_j^m\| < \varepsilon \) and 0 otherwise.

The correlation dimension is given by:

\[ \nu_m = \lim_{\varepsilon \to 0} \frac{C_m(\varepsilon)}{\log(1/\varepsilon)} \]  

4

If there is determinism in the dynamics of the system,
the correlation dimension will increase with the increase of
the embedding dimension and finally it will saturate to a
certain value. If the system is random the correlation
dimension will never saturate.

To determine the largest Lyapunov exponent we apply
Wolf algorithm. If we consider two distinct moments in
time \( T_0 \) and \( T \) and of two diverging trajectories in the state
space, we can measure the rate of divergence as the
magnitude of deformation of a initial sphere containing the
trajectories, towards an ellipsoid (Figure 2).

The maximum LE is given by:

\[ \lambda_{\text{max}} = \lim_{t \to \infty} \lim_{\varepsilon(0) \to 0} \left( -\frac{1}{t} \log_2 \left( \frac{\varepsilon(t)}{\varepsilon(0)} \right) \right) \]  

5
Where $\varepsilon_i(t)$ is the length of the principal axis of the ellipsoid at time $t$.

IV. RESULTS

The existence of an initially strong autocorrelation indicates the fact that the time series is not random (Figure 3).

The attractor was embedded in the reconstructed state space using $\tau=10$ as the first minimum of mutual information function (Figure 4) and $m=8$ the embedding dimension for which the percentage of FNN drops below 0.1% (Figure 5).

According to the Takens’ theorem, the attractor’s dimension is around 3.5 which indicates the presence of a low dimensional attractor. The correlation dimension (Figure 6) saturates around 3.2, confirming the results obtained using FNN method.

The phase space plot reveals the existence of an attractor occupying a bounded region of this space.
If the system is chaotic, the rate of the information loss (the speed at which the deterministic component vanishes) is given by LE.

The maximum LE was determined to be 0.071 indicating that nearby trajectories are diverging. Taking into account that the correlation dimension saturates we may say that we have proven the existence of a low dimension chaotic attractor.

V. CONCLUSIONS

The paper presents some of the tools of nonlinear time series analysis applied in order to investigate the possible existence of chaos in the system that produced the time series. After we preprocessed the data set (the noise was removed) we tried to establish if the principal conditions for the existence of chaotic dynamics were met. We have determined that the correlation dimension saturates and the maximum Lyapunov exponent is positive. These findings are important, proving that short term forecasting is possible.

VI. REFERENCES


Sorin Vlad received both and BSc degrees in Computer Science and Engineering from “Stefan cel Mare” University of Suceava, Romania. He is now a PhD student at the same university. His research interests include Nonlinear Time Series Analysis and Neural Networks.

Ştefan Gheorghe Pentiu is a professor at the “Stefan cel Mare” University of Suceava, Faculty of Electrical Engineering and Computer Science. His research interests include Pattern Recognition, Parallel and Distributed Systems.