Chaos Theory for Forecasting

WORD COUNT	2534	TIME SUBMITTED	20-MAY-2021 02:14PM
		PAPER ID	72354147

Chaos theory for forecasting

D.I. Kusumastuti
The University of Lampung, Bandar Lampung, Indonesia

R.K. Price
IHE, Delft, The Netherlands

ABSTRACT : Given a time series for runoff in a river there is often the need to make short term forecasts of flows, particularly if they can lead to possible flooding. The use of chaos theory in different sectors to analyse apparently chaotic systems has provided a set of useful tools. In each case the forecast from chaos theory works better than the other techniques, such as artificial neural network based on extrapolating one or more time steps ahead and nearest neighbour techniques. The decay rate of the forecast is developed in terms of the structure of the time series.

1 INTRODUCTION

This case study was carried out at a number of sites in Tulang Bawang basin in Indonesia. About 20 years of daily runoff data were available from 5 flow gauges in Tulang Bawang basin. There is some concern about the timing of the daily measurements in the Tulang Bawang basin, so an approach was developed using an artificial neural network to generate the (daily) time series from daily rainfall data. The multilayer perceptron method was used to train the network in order to fill in missing daily runoff data.

The daily runoff time series data were analysed for an underlying chaotic. A systematic methodology for analysing the measurement data had to be developed to indicate the presence of chaos in the data set.

2 CHAOS THEORY

Chaos comprises a class of signals intermediate between regular sinusoidal or quasiperiodic motions and unpredictable, truly stochastic behaviour. It has long been seen as a form of 'noise' because the tools for its analysis were couched in a language tuned to linear processes. Chaos, as a property of orbits x(t), manifest itself as complex time traces with continuous, broadband Fourier spectra, nonperiodic motion, and exponential sensitivity to small changes in the orbit. With conventional linear tools such as Fourier transform, chaos looks like 'noise', but chaos has structure in an appropriate state or phase space. This structure means there are numerous potential engineering applications to sources of chaotic time series which can take advantage of the structure to predict and control those sources (Abarbanel, 1996).

One set of data used in the process of developing the models above consisted of recorded runoff from Rantau Jangkung flow gauge in the period 1974-1993. An easy tool for identifying irregularities in observed signals is the Fourier Power Spectrum as shown in Figure 1. A broadband and continuous spectrum indicates that the signal has possibly originated from a chaotic system. However, the Fourier power spectrum is not itself a true invariant because new frequencies will be introduced with nonlinear changes to the coordinate system, and multi-harmonic outputs do not necessarily indicate that the system is chaotic. The Fourier power spectrum only provides an indication that the system might have a chaotic origin because systems with large degrees of freedom may generate similar power spectra.

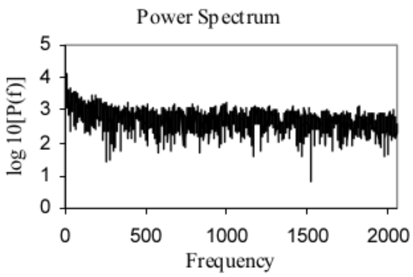

Figure 1: Power spectrum of the runoff in Rantau Jangkung flow gauge. The spectrum is broadband and continuous, indicating a non-periodic signal.

An auto-correlation of a periodic signal produces a periodic function. For a chaotic or a random signal the autocorrelation function will approach zero rapidly. This tool is a good indicator of whether the system is chaotic or random in nature.

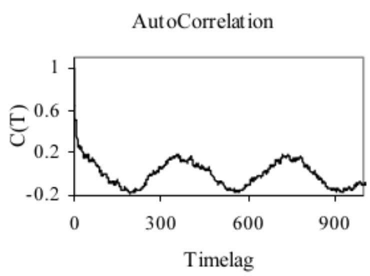

Figure 2: The autocorrelation function plotted against time lags for the runoff in Rantau Jangkung flow gauge.

2.1 Reconstruction of phase space

To go from scalar or univariate observations to the multivariate state or phase space which is required for chaotic motions, it is necessary to construct the data vectors y(n):

$$y(n) = [s(n), s[n+T], \ldots, s(n+T(d-1))] \qquad (1)$$

composed simply of time lags of the observation at time $n\tau_s$. These y(n) replace the scalar data measurements s(n) with data vectors in an Euclidean d-dimensional space in which the invariant aspects of the sequence of points x(n) are captured with no loss of information about the properties of the original system. The appropriate time lag $T\tau_s$ and the dimension d are the central issues oinreconstructing the phase space.

2.1.1 Average mutual information

.

The mutual information between measurement a_i drawn from a set $A = \{a_i\}$ and measurement b_j drawn from a set $B = \{b_j\}$ is the amount to be learned by the measurement of a_i from the measurement of b_j. The average over all measurements, called the average mutual information between measurements A and B, is

$$I_{AB} = \sum_{a_i, b_j} P_{AB}(a_i, b_j) \log_2 \left[\frac{P_{AB}(a_i, b_j}{P_A(a_i) P_B(b_j)} \right] \qquad (2)$$

where $P_{AB}(a,b)$ is the joint probability density for measurements A and B resulting in values a and b. $P_A(a)$ and $P_B(b)$ are the individual probability densities for the measurements of A and of B.

The choice of values for the time lag is accomplished by asking when the nonlinear correlation function, called average mutual information, has its first minimum. This function gives the amount of information, in bits, learned about $x(t + T\tau_S)$, at time $\tau_T = T\tau_S$, from the measurement of $x(t + (T-1)\tau_S)$. The integer multiple T of the sampling time is the goal. The average mutual information among measurements of the runoff at Rantau Jangkung flow gauge is seen in Figure 3. Here it is observed that the first minimum is at T = 15.

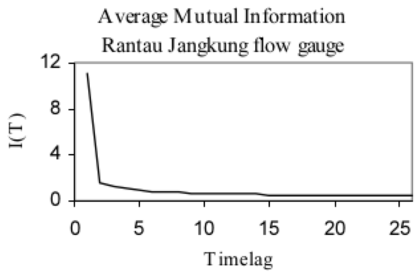

Figure 3: The average mutual information I(T) for runoff in Rantau Jangkung flow gauge.

2.1.2 Global Dimension

The global embedding dimension, d_E , is the lowest dimension which unfolds the attractor so that none of these overlaps remain. d_E is an integer. If the dimension of the attractor defined by orbits is d_A, then the attractor can be unfolded in an integer dimensional space of dimension d_E where $d_E > 2d_A$.

If the signal is contaminated, however, it may be that the contamination will so dominate the signal of interest that we see instead the dimension required to unfold the contamination. If the contamination has a very high dimension, such as would be anticipated for noise, then it cannot be seen that the percentage of false nearest neighbours will drop anywhere near to zero in any dimension. A plot of the percentage of false neighbours for runoff in Rantau Jangkung flow gauge is shown in Figure 4.

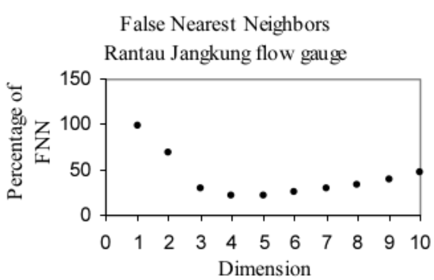

Figure 4: Global false nearest neighbour; a time delay T=15 is used

From this it can be seen that the percentage of false nearest neighbours does not drop to zero in any dimension, and this indicates the presence of a high dimensional signal in the data. The high residual level of global false nearest neighbours is about 21 percent where the dimension is 4. We interpret this level of high dimensional signal as arising from noise.

2.1.3 Lyapunov exponents and correlation dimension

The stability of the system is determined by the Lyapunov exponents, which tell us how small changes in the orbit will grow or shrink in time. On taverage, the Lyapunov exponents indicate how well a prediction could be made of the evolution of the system L steps ahead of the present location. Since chaotic systems are extremely sensitive to initial conditions, Lyapunov exponents provide a good test for chaos.

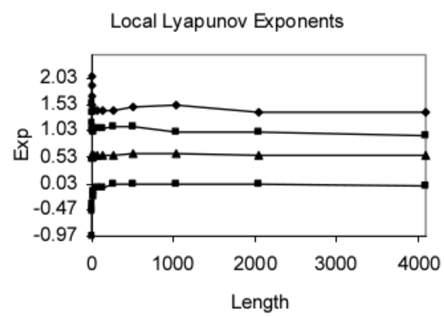

Figure 5: Average local Lyapunov for runoff in Rantau Jangkung.

Figure 5 shows four Lyapuov exponents for runoff in Rantau Jangkung flow gauge. Three positive exponents suggest that the trajectories are diverging exponentially, which indicate that the system is chaotic. The negative exponent means a dissipative mechanism exist within the system. The presence of four exponents also indicates that there are four degrees of freedom, which implies that the behaviour of the system could be described by four differential equations.

The sum of all four Lyapunov exponents is, in this case, positive. This indicates that overall the system is diverging, so that it can be expected that data may be polluted by a large degree noise.

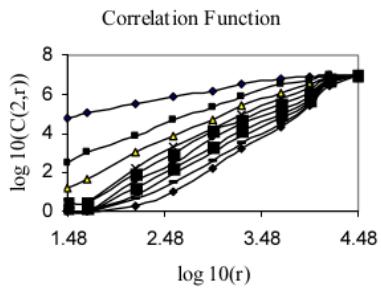

Figure 6: Correlation dimension for runoff in Rantau Jangkung flow.

The correlation dimension of different radii r can be estimated from Figure 6. The fractal dimension estimated from the slopes of the curves is in the range 0.9 – 2.2. This indicates a non-integer, fractal dimension for the attractor and is an indication that chaotic dynamics may present.

From the discussion above it can be concluded that runoff in Rantau Jangkung flow gauge exhibits chaotic behaviour. This, however, does not agree with the global false nearest neighbour curve, which should drop to zero and would remain zero from then on. This indicates there is significant random contribution in the data.

There is quite a bit of information in the local exponents, since they tell us how well we can predict the evolution of the system T step ahead. The prediction will work accurately within the limits of prediction dictated by the largest Lyapunov exponent λ_1.When we try to predict beyond the instability horizon, that is, for times much greater than $\frac{\tau_s}{\lambda_1}$, our prediction should rapidly lose accuracy.

3 CHAOS PREDICTION

The phase space structure that has been built up out of the $y(n)$ can be used to provide effective models of the dynamics which enable the prediction of the evolution of any new point in the phase space which has been observed. Since it has been seen how points in a neighbourhood evolve into points in the following neighbourhood, the next step is to provide an appropriate interpolation scheme that allows us to say that any new point evolves more or less as the points in its neighbourhood.

The Chaos prediction is a purely predictive model. It does not use rainfall as a forcing term, but generates predictions on the basis of the observed runoffs only. The first ten years of daily data from 1974 – 1983 were used for training, while the second ten years were used for testing. The Chaos model was built for the horizon of T= 1 and T = 2 using embedding dimension d = 4, time sampling rate τ = 2. Performance of the testing data is presented in Figure 7.

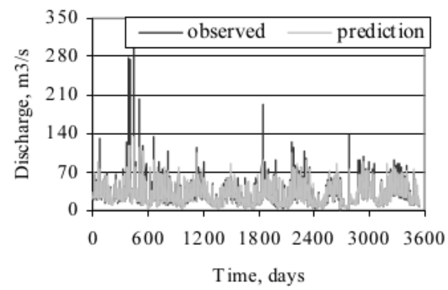

Figure 7. Results of runoff prediction in Rantau Jangkung (T=2).

The performance of the model shows good prediction for lead time T =1 with coefficient of goodness of fit of 0.9116. This shows that chaos can be used to give precise short-term forecasting if the time series is indeed chaotic and low-dimensional. Comparing the results of forecasting of river discharges by ANN, Genetic Algorithm, and Chaos Theory showed that even for an horizon of T = 2, results are comparable with that of neural network for forecast horizon of T = 1 (Babovic et al., 1999).

In units of τ_s the largest Lyapunov exponent from Figure 5 is approximately $\lambda_1 \approx 1.4$, so we should be able to make accurate predictions for one or two steps beyond any starting location on the attractor.

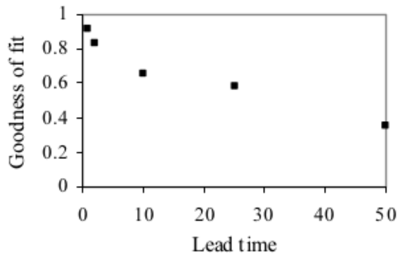

Figure 8: The decay rate of the forecast.

Figure 8 exhibits the attempt to predict for 1, 2, 10, 25, and 50 steps using precisely the same methods and the same learning data set. It is clear that we have gone beyond the predictability horizon by pushing so far, since we are dealing with a chaotic evolution which is unstable everywhere in state space.

4 FORECASTING BY ANN AND NEAREST NEIGHBOUR

4.1 Forecasting by ANN

The most widely used neural network architecture for modeling water resources systems is the MLP network with the error back-propagation algorithm as learning rule. However, a major limitation of the MLP is that it can only learn an input-output mapping that is static. These networks are well suited for the prediction of outputs from inputs where the outputs do not affect any of the other input patterns. However, it is questionable how far MLP can be used for forecasting in time.

The network used for prediction is a three layer-ANN with five inputs, eight hidden nodes, and one output node. The network was trained using backpropagation algorithm. The network is same as 'the usual' MLP network, except the desired output dynamically changes as time prediction changing. In practice, it can be simply done by shifting the output one step for one step higher time level.

For one step prediction, it can be considered that MLP gives a good result with the coefficient of efficiency 0.7339 and 0.6199 for training and verification respectively. However, when the networks were trained for higher time level prediction these values decrease to be unacceptable.

4.2. Forecasting by nearest neighbour

The nearest neighbour algorithm may be stated briefly as follows. Training set patterns are first plotted in multi-dimensional feature space, and then test patterns are taken one at a time and classified according to which training set pattern is the nearest in feature space (Dudani, S.A., 1990).

For every data vector in training set the value of precedent runoff is known. With the new vector we want to forecast runoff for one step ahead.

5. CONCLUSION

Chaos has structure in an appropriate state or phase space in which that structure can be used to predict the source of the chaos. A systematic methodology for analysing the measurement data depends on a reconstruction of the phase space from scalar observation, choosing time delays, deducing the false nearest neighbour for determining the size of the reconstructed space, and modelling of the observed systems. The experimental results indicated the presence of chaos in the data set, but there remains uncertainty that runoff in Tulang Bawang river basin actually exhibits chaotic behaviour.

The prediction by Chaos Theory worked remarkably accurately within the limits of the prediction prescribed by the largest lyapunov exponent λ_1. With the set of data the predictions are accurate with horizons 1 and 2. However, predictions beyond this horizon lead rapidly to loss of accuracy. The experiments with 10, 25, and 50 steps ahead showed how the accuracy degrades

In forecasting, Chaos Theory shows its superiority compared with ANN and nearest neighbour (NN). Results displayed in Table 1, for an experiment in Ranrau Jangkung flow gauge, show that even for a lead time T=4, results are comparable with that of neural network and nearest neighbour for forecast lead time of T=1.

Table 1. The comparison of the results

Lead Time	Chaos Theory	ANN		NN
		Train	Test	
1	0.912	0.734	0.619	0.752
2	0.834	0.636	0.532	0.682
3	0.781	0.567	0.462	0.622
4	0.774	0.512	0.416	0.560
5	0.713	0.471	0.358	0.499

REFERENCES

Abarbanel, H.D.I., "Analysis of Observed Chaotic Data," Springer-Verlag New York, Inc., New York, Inc., New York, 1996.

Babovic, V. and Keijzer, J., "Forecasting of river discharges in the presence of chaos and noise," *NATO Advanced Research Workshop on Coping with Floods : Lessons Learned from Recent Experience*, 1999

Dudani, S.A., "The distance-weighted k-nearest neighbor rule", *Nearest Neighbor Pattern Classification Techniques*, IEEE Computer Society Press, Los Alamitos, 1990.

Chaos Theory for Forecasting

ORIGINALITY REPORT

3%

SIMILARITY INDEX

PRIMARY SOURCES

1	Michael Siek. "Multivariate chaotic models vs neural networks in predicting storm surge dynamics", 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 06/2008 Crossref	23 words — 1%
2	nora.nerc.ac.uk Internet	23 words — 1%
3	A. Albostan, B. Barutcu, B. Onoz. "A non-linear time series prediction method for missing daily flow rate data of Middle Firat Catchment", Energy Procedia, 2011 Crossref	11 words — < 1%
4	citeseerx.ist.psu.edu Internet	10 words — < 1%
5	Marek Kubalčík, Vladimír Bobál. "ADAPTIVE CONTROL OF THREE - TANK - SYSTEM: POLYNOMIAL APPROACH", IFAC Proceedings Volumes, 2006 Crossref	8 words — < 1%
6	eprints.nottingham.ac.uk Internet	7 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	ON		