# CHAPTER 4. TESTING FOR CHAOS IN TRAFFIC FLOW DYNAMICS

It is argued that testing for chaos is more artistic than scientific and that no recipe will guarantee success for every case (Sprott, 2003). Nonetheless, this chapter attempts to develop a parsimony procedure to test if chaotic phenomena exist in a traffic flow dynamics. We undertake a comprehensive comparison of promising plots and statistics between the observed freeway traffic flow data and their surrogates. The most crucial indexes are selected to develop the parsimony procedure. We also utilize some well-known time series data generators to validate the proposed procedure and further apply it to test for the chaotic of traffic flows at different sites. This chapter is organized as follows: In section 4.1, introduces the original and surrogate data. In section 4.2, an empirical study is carried out using the minute-flow time series data drawn from the I-35 Freeway in Minneapolis, Minnesota. A comparison is then made between the original traffic flow data and their surrogates. Section 4.3 selects the most crucial indexes to develop a parsimony testing procedure, which is further validated by well-known time series generators.

## 4.1 Original and Surrogate Data

Our empirical one-minute traffic flow time series data are directly drawn from 16 detector stations of the United States I-35 Freeway in Minneapolis, Minnesota. Averages of the lane-specific flow counts are accumulated over one-minute period. At each station a total of 1,780 minute-flow samples are used, representing ten workdays' morning peak hours from 6 am to 9 am each day (Note: 20 samples are missing on the last day at some stations, thus we only take 1,780 samples). The average lane-flow rates for these 16 stations range from 16.8 to 33.8 vehicles per minute, or equivalently, with average headways from 3.57 seconds (a moderate flow) to 1.78 seconds (a saturated, near capacity, flow).

As pointed out by Sprott (2003), a stochastic system with a non-uniform power

spectrum can masquerade for chaos. Therefore, we must test our conclusions about whether an original time series is chaotic by further applying the test to the surrogate data, which are designed to mimic the statistical properties of the original data, but with the determinism removed. The surrogate data are generated by randomly shuffling the original data. The shuffling can be done quickly by stepping through the time series, swapping each values with one chosen randomly from anywhere in the series. On average, each point being moved twice will essentially guarantee randomness. This method does require keeping the whole time series in memory. While shuffling the sequences will preserve the same probability distribution as the original data, the surrogates do not preserve the same power spectrum and correlation function (Theiler, et al. 1992). In other words, the surrogates are not chaotic.

To generate the surrogate time series $Y_n$ with the same power spectrum as the original time series $X_n$, we use the following equations to find $S_m$ and then to construct a surrogate series $Y_n$ with the same Fourier amplitudes but with random phases (Osborne, *et al*. 1986).

$$S_m = G(0) + G(k)\cos(\frac{2\pi mk}{N}) + 2\sum_{k=1}^{K} G(k)\cos(\frac{2\pi mk}{N}) \quad (4\text{-}1)$$

$$Y_n = \frac{a_0}{2} + \sum_{m=1}^{N/2} \sqrt{S_m}\sin 2\pi(mn/N + r_m) \quad (4\text{-}2)$$

where *G(k)* is autocorrelation function, *K* is the maximum of *k*, which usually appears to be about *N/4*. $r_m$ are *N/2* uniform random numbers chosen from $0 \le r_m < 1$. *N* is the number of data. The surrogate data for the 16 stations are generated by the above equations, eight of which are used to establish the parsimony procedure and the remaining eight stations are reserved to examine the applicability of our proposed parsimony procedure.

## 4.2 Empirical Testing Results

By visualizing all the geometric plots for the eight stations, we find that some plots have not displayed apparent difference between the original and surrogate data, but some others have noticeable distinction. Since the details of geometric plots for the eight stations are tediously long, the paper only demonstrates the details at station

50, in which the average lane-flow is 25.6 vehicles per minute (or, average headway is 2.34 seconds).

The one-dimensional graph of data in Figure 4-1(a) verifies the integrity of the original time series, which ranges reasonably within a bounded interval between zero and around 35 vehicles per minute per lane (equivalently, the lane capacity is nearly 2,200 vehicles per hour), suggesting that there are no outrageously silly mistakes, formatting errors or missing points of the observed traffic data. We notice that the flow rates in the seventh day are systematically lower than the other days. It could be due to bad weather (e.g., raining) or a long-duration incident (such as road construction work). The graph of data in Figure 4-1(b) presents almost complete randomness, showing that the surrogate data has no deterministic structure at all. The trajectories of surrogate time series in Figures 4-2(b) and 4-3(b) scatter a bit more uniformly in the plane or space than the original time series in Figures 4-2(a) and 4-3 (a), implying that the original data is not as random as the surrogate data. However, these two- and three-dimensional graphs of data do not differ significantly. The return map in Figure 4-4(a) appears a little more than that in Figure 4-4(b) in the form of a strange attractor (fractal structure, an evidence of chaos), implying that the original data is likely chaotic. However, both Figures 4-4(a) and 4-4(b) do not reveal noticeable dissimilarity. The phase-space plots of the original data in Figures 4-5(a) and 4-6(a) do not differ significantly from surrogate data in Figures 4-5 (b) and 4-6 (b), either. The Poincare movies in Figures 4-7(a) and 4-8(a) show the repeated stretching and folding of the trajectories, indicating that the data is likely chaotic. Figures 4-7(b) and 4-8(b) do not reveal such property but their Poincare sections are not significantly different from the original data. In sum, the geometric plots from Figure 4-2 to Figure 4-8 do not differ conspicuously between the original and surrogate data; thus, these plots should not be good indexes for analyzing the chaotic structures.

Theoretically, the IFS clumpiness map is a good indicator of determinism. If the data is white noise it should fill the screen uniformly. The IFS clumpiness map in Figure 4-9(a) presents localized clumps, strongly inferring that the original data is not white noise. In contrast, the IFS clumpiness plot in Figure 4-9(b) fills up the screen uniformly, firmly indicating that the surrogate data is white noise. The same

probability distribution for both original and surrogate data in Figures 4-10(a) and 4-10(b) confirms that the distribution of surrogate time series has been preserved even though we have shuffled the order form the original data. Note that the probability distribution does not form a simple histogram with sharp edges, strongly implying that the traffic flow dynamics is neither periodic nor quasi-periodic. Figure 4-11(a) visibly shows that the correlation function does not drop abruptly to zero as Figure 4-11(b) does, strongly indicating that the original data is not random but the surrogate data is. Figure 4-12(a) shows that the data gives rise to broad spectra without producing a few dominant peaks, also clearly indicating that the original data is neither periodic nor quasi-periodic. In contrast, the copious sharp peaks in Figure 4-12(b) provide strong evidence of randomness for the surrogate data. In sum, except for Figures 4-10(a) and 4-11(b) that are supposed to be the same, the geometric plots in Figures 4-9 through 4-12 are apparently dissimilar between the original and surrogate data; thus, these plots could be good indexes for analyzing the chaotic structure.

We further calculate all of the aforementioned statistics for the same eight stations and the results are reported in Table 4-1. The largest Lyapunov exponent (LE) for both original and surrogate data are positive, which has definitely ruled out both time series being periodic or fixed point. The embedding dimension (ED), capacity dimension (CAD) and correlation dimension (COD) of the original data are somewhat smaller than those of the surrogate data, implying that the original data is less random than the surrogates. A lower dimensionality of such indexes often suggests that the data is not random. The original data has smaller Kolmogorov entropy (KE) than the surrogates, implying that the original data is less disorder than the surrogates. Except for the largest Lyapunov exponent that is the most critical index for ruling out the time series is likely to become eventually stable fixed points (LE<0) or periodic (LE=0), all of the above indexes in effect do not differ largely between the original and surrogate data, thus will not be used for analyzing the chaotic structure.

The delay time (DT), a proxy of correlation function, also reaches the same conclusion as by Figure 4-11. The surrogate data has a very small DT value while the original data exhibits a large DT value, strongly suggesting that the surrogate data is random, while the original data is not. Hurst exponent (HE) is different from 0.5 at

each station, implying that the original data is not Brownian noise (random walk). Note that HE equal to zero for the surrogates in fact represents a pink noise, very similar to the white noise whose exact HE value is –0.5 (Sprott, 2003). The relative complexity (LZC) is equal to 1.0 for the surrogates, indicating that the time series is completely randomized (has the highest complexity). However, all the original data also have rather large LZC values (0.6~0.8), implying that the minute-flow time series is very oscillatory so that it is difficult to model or predict.

Table 4-1 Statistics of minute-flow time series for eight stations

| Station No. | Average lane-flow (veh./min) | Data | DT | ED | COD | LE | KE | HE | LZC | CAD |
|---|---|---|---|---|---|---|---|---|---|---|
| 49 | 21.9 | original | 12.1 | 3.0 | 3.2 | 0.7 | 0.2 | 0.2 | 0.8 | 2.4 |
| | | surrogate | 0.7 | 4.0 | 3.9 | 0.5 | 0.3 | 0.0 | 1.0 | 3.6 |
| 48 | 23.9 | original | 12.4 | 3.0 | 3.2 | 0.7 | 0.2 | 0.2 | 0.8 | 2.5 |
| | | surrogate | 0.6 | 4.0 | 3.9 | 0.6 | 0.4 | 0.0 | 1.0 | 3.5 |
| 50 | 25.6 | original | 13.9 | 3.0 | 3.2 | 0.7 | 0.4 | 0.2 | 0.7 | 2.5 |
| | | surrogate | 0.6 | 5.0 | 4.5 | 0.5 | 0.4 | 0.0 | 1.0 | 4.1 |
| 39 | 27.8 | original | 11.0 | 3.0 | 3.2 | 0.7 | 0.2 | 0.1 | 0.8 | 2.0 |
| | | surrogate | 0.6 | 4.0 | 3.9 | 0.4 | 0.4 | 0.0 | 1.0 | 2.9 |
| 42 | 31.9 | original | 11.9 | 4.0 | 3.9 | 0.4 | 0.3 | 0.1 | 0.8 | 2.5 |
| | | surrogate | 0.6 | 5.0 | 4.6 | 0.3 | 0.5 | 0.0 | 1.0 | 3.3 |
| 51 | 32.5 | original | 14.0 | 3.0 | 3.2 | 1.1 | 0.3 | 0.2 | 0.7 | 2.4 |
| | | surrogate | 0.6 | 6.0 | 5.0 | 0.5 | 0.7 | 0.0 | 1.0 | 4.1 |
| 54 | 32.8 | original | 16.1 | 4.0 | 3.8 | 0.4 | 0.5 | 0.3 | 0.7 | 2.9 |
| | | surrogate | 0.6 | 4.0 | 3.9 | 0.6 | 0.4 | 0.0 | 1.0 | 3.6 |
| 55 | 33.8 | original | 16.2 | 4.0 | 3.8 | 0.4 | 0.3 | 0.3 | 0.7 | 3.0 |
| | | surrogate | 0.6 | 4.0 | 3.9 | 0.6 | 0.5 | 0.0 | 1.0 | 4.2 |

Based on the comparative results of geometric plots and statistics between the original and surrogate data, we are confident that the observed freeway minute-flow dynamics is not periodic, quasi-periodic, or eventually fixed points. It is not white noise or random. As a consequence, we can conclude that the minute-flow dynamics exhibits chaotic structures.

## 4.3Development of Parsimony Procedure

4.3.1 Proposed testing procedure

We attempt to establish a "parsimony" testing procedure by utilizing as few geometric plots or statistics as possible. Only the most crucial or significant indexes will be included in the procedure. Figure 4-13 depicts the procedure, which is briefly narrated as follows:

Step 1. Examine the largest Lyapunov exponent (LE) for the original data

If the LE is negative, the time series will converge toward a stable sink (called equilibrium fixed points). If the LE is zero, the time series is periodic in the sense that the trajectories will converge to a period-k sink (k is greater or equal to 2). If the LE is positive, the time series can be quasi-periodic (with multiple incommensurable periods) or chaotic (with determinism) or stochastic (with random noise), then we go to step 2.

Step 2. Examine the power spectrum of the original data

If the power law spectrum is narrow and has only few (two or three) dominant sharp peaks, it must be quasi-periodic. In case that it is a broadband spectrum, it can be chaotic or stochastic, then we go further to step 3.

Step 3. Compare the IFS clumpiness between the original and surrogate data

If the IFS clumpiness map of the original data visually differs from the surrogate, then we have evidence (but not proof!) that the time series is not stochastic (Smith 1992a). In this case, we can probably say that the original time series exhibits chaotic structures.

In short, the first step is to rule out the fixed points and periodic data. The second step further rules out the quasi-periodic data. The final step further makes distinction between chaotic and stochastic time series.

4.3.2 Validations

To validate how well our proposed parsimony testing procedure work, we utilize nine well-known time series data generators: FIX.DAT is fixed-point time series generated by eq. (4-3). FEIGEN.DAT (The iterates of this map form a

one-dimensional Cantor set at the transition between periodic and chaotic behavior) and SINE.DAT are periodic time series generated by eqs. (4-4) and (4-5). TWOSINE.DAT and THREESIN.DAT are quasi-periodic time series generated by eqs. (4-6) and (4-7). NOISE.DAT and RANDOM.DAT are stochastic time series generated by eqs. (3-3) and (4-8). HENON.DAT and LORENZXZ.DAT are chaotic time series generated by eqs. (3-1) and (3-2). We produce 2,000 iterates for each of the nine time series generators and then examine the largest Lyapunov exponent in the first step, the power spectrum in the second step and the IFS clumpiness in the final step proposed in the above-mentioned parsimony procedure.

$X_n$=constant (4-3)

$X_{n+1}=3.5699456X_n(1-X_n)$ (4-4)

Sin (t/10) (4-5)

where t is an integer value from 0 to 1999.

Sin (t/2)+cos(gt/2) (4-6)

where t is an integer value from 0 to 1999,g=$(\sqrt{5}-1)/2$

Sin (t/2)+cos(gt/2)+sin(ht/2+$\pi$/4) (4-7)

where t is an integer value from 0 to 1999, g=$(\sqrt{5}-1)/2$

$$X=\sqrt{-2\ln r_1}\sin 2\pi r_2 \quad (4\text{-}8)$$

Table 4-2 presents the largest Lyapunov exponents of these nine generators. Figure 4-14 displays the power spectra, and Figure 4-15 displays the IFS clumpiness. From Table 4-2, the first step has successfully identified the FIX.DAT (LE<0) as an equilibrium fixed-point time series and SINE.DAT (LE=0) and FEIGEN.DAT (LE=0) as periodic time series. Therefore we go to step two by examining the power spectra for the rest of the tested data. We find that only TWOSINE.DAT (Figure 4-14(a)) and THREESIN.DAT (Figure 4-14(b)) have spectra with few dominant peaks, which are evidences for quasi-periodic time series. Finally, we go to step three to further

distinguish chaotic from stochastic time series by comparing the IFS clumpiness maps between original and surrogate data. We find that the original data of NOISE.DAT and RANDOM.DAT (Figure 4-15(a) and (c)) do not differ from the surrogate (Figure 4-15 (b) and (d)), indicating that NOISE.DAT and RANDOM.DAT are stochastic data. By contrast, the original data (Figure 4-15(e) and (g)) differ from the surrogate (Figure 4-15(f) and (h)), supporting that HENON.DAT and LORENZXZ.DAT are chaotic data. In other words, these known time series data generators validate our proposed parsimony testing procedure.

Table 4-2 The parsimony procedure validated by known time series generators

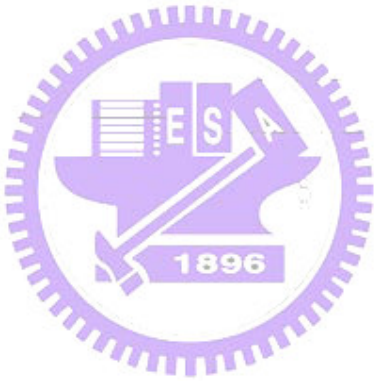
| Data Generator | Known Properties | Largest Lyapunov Exponent (LE) |
|---|---|---|
| FIX.DAT | Fixed point | -0.1 |
| SINE.DAT | Periodic | 0.0 |
| FEIGEN.DAT | Periodic | 0.0 |
| TWOSINE.DAT | Quasi-periodic | 0.5 |
| THREESIN.DAT | Quasi-periodic | 0.5 |
| NOISE.DAT | Stochastic | 0.5 |
| RANDOM.DAT | Stochastic | 0.7 |
| LORENZXZ.DAT | Chaotic | 0.1 |
| HENON.DAT | Chaotic | 0.6 |

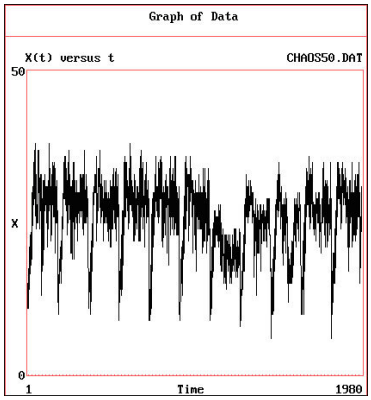
### 4.3.3 Applications

We apply our proposed parsimony procedure to further test the remaining ten stations on the I-35 Freeway whether or not chaotic structures also exhibit in the minute-flow dynamics. Table 4-3 reports the results. The LE value for each station is positive, ruling out the traffic dynamics is equilibrium fixed points or period. The power spectrum for each station is broad (station 32, for example, Figure 4-16), further ruling out the flow time series quasi-period. Finally, the IFS clumpiness for each station (station 32, for example, Figure 4-17) reveals an obvious difference between the original and surrogate data, further ruling out the data stochastic. Thus, we conclude that all the remaining eight stations also exhibit chaotic structures in the nature of their one-minute flow dynamics.

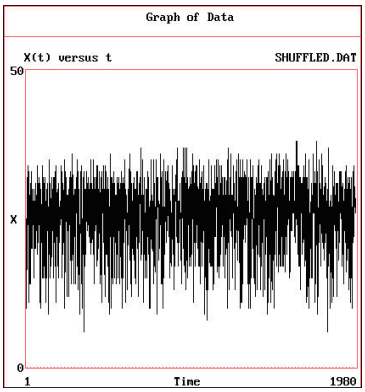
Table 4-3 Applications of the parsimony procedure

| Station No. | Average lane-flow (veh./min) | Largest Lyapunov Exponent (LE) |
|---|---|---|
| 32 | 16.8 | 0.6 |
| 45 | 23.7 | 0.7 |
| 44 | 24.8 | 0.7 |
| 43 | 28.9 | 0.7 |
| 41 | 29.9 | 0.7 |
| 52 | 31.9 | 0.3 |
| 53 | 32.5 | 0.4 |
| 56 | 33.2 | 0.4 |



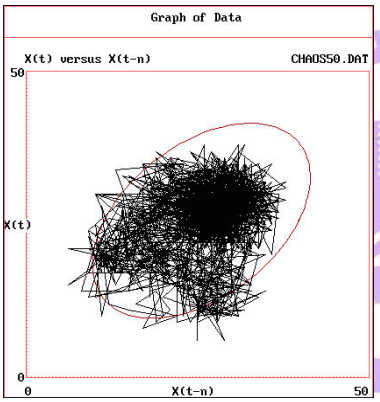



(a) Original data (station 50)

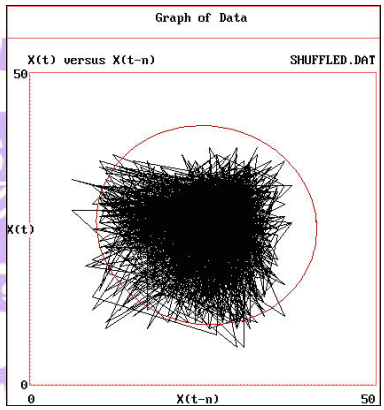



(b) Surrogate data (station 50)

Figure 4-1 One-dimensional state-space plots

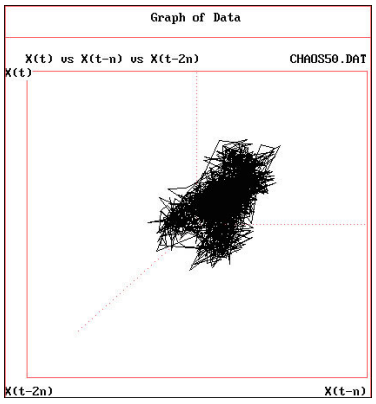



(a) Original data (station 50)

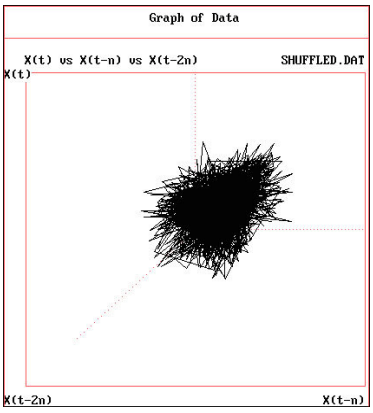



(b) Surrogate data (station 50)

Figure 4-2 Two-dimensional state-space plots

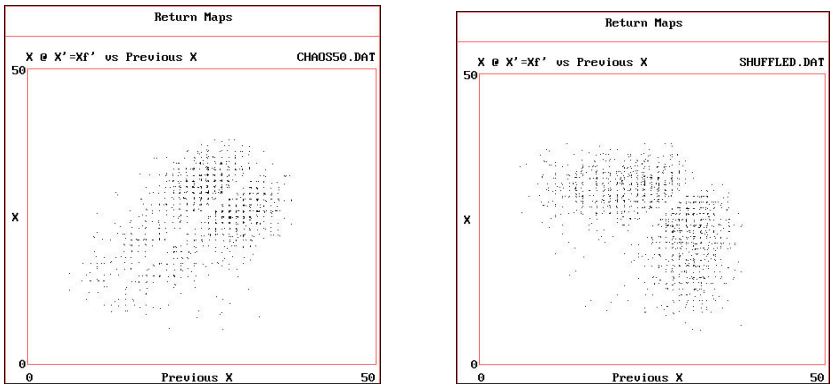



(a) Original data (station 50)



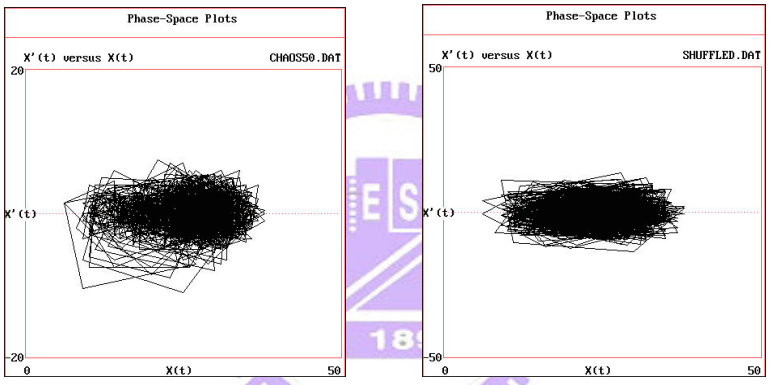

(b) Surrogate data (station 50)

Figure 4-3 Three-dimensional state-space plots



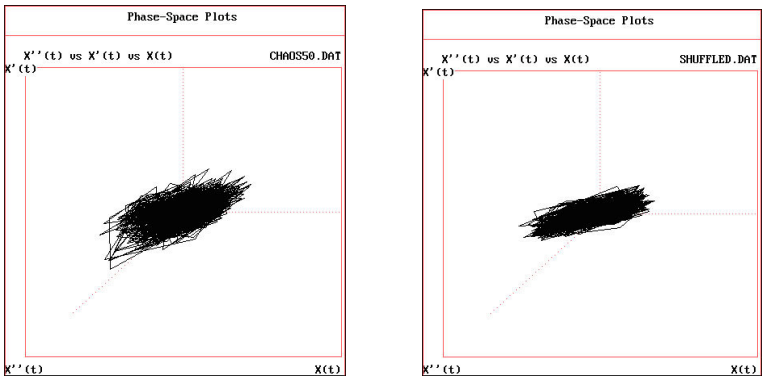

(a) Original data (station 50) (b) Surrogate data (station 50)

Figure 4-4 Return maps



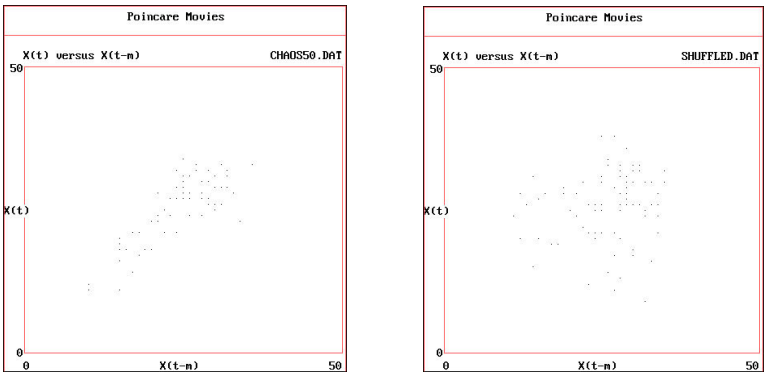

(a) Original data (station 50) (b) Surrogate data (station 50)

Figure 4-5 Two-dimensional phase-space plots



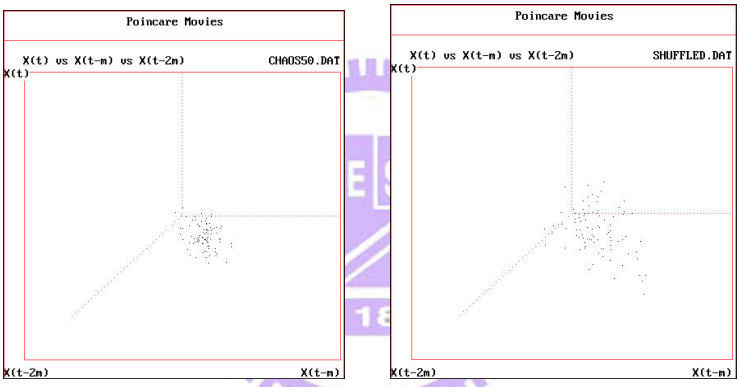

(a) Original data (station 50) (b) Surrogate data (station 50)

Figure 4-6 Three-dimensional phase-space plots



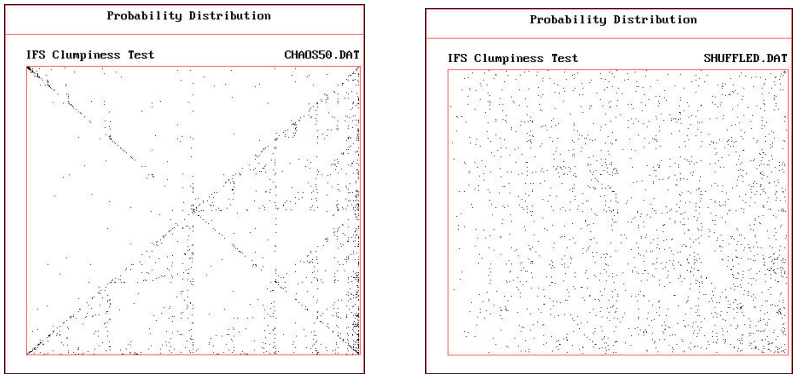

(a) Original data (station 50) (b) Surrogate data (station 50)

Figure 4-7 Two-dimensional Poincare' movies



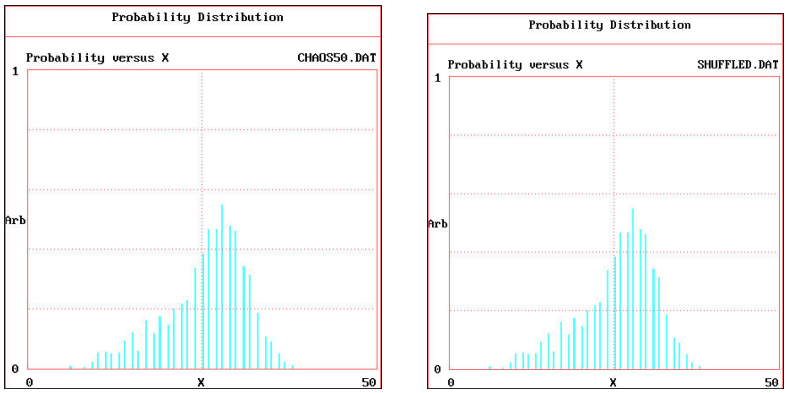

(a) Original data (station 50) (b) Surrogate data (station 50)

Figure 4-8 Three-dimensional Poincare' movies



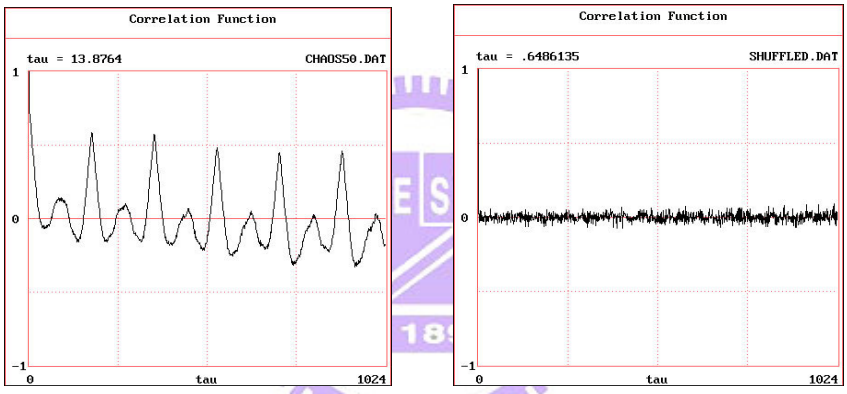

(a) Original data (station 50) (b) Surrogate data (station 50)

Figure 4-9 IFS clumpiness maps



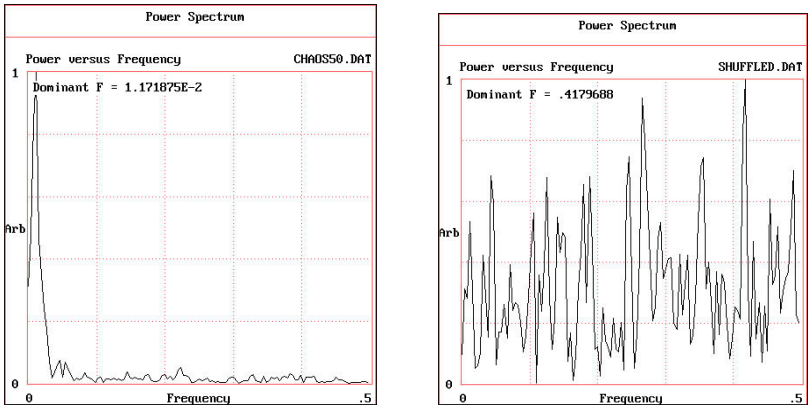

(a) Original data (station 50) (b) Surrogate data (station 50)

Figure 4-10 Probability distributions




(a) Original data (station 50) (b) Surrogate data (station 50)

Figure 4-11 Correlation function




(a) Original data (station 50) (b) Surrogate data (station 50)

Figure 4-12 Power spectrum

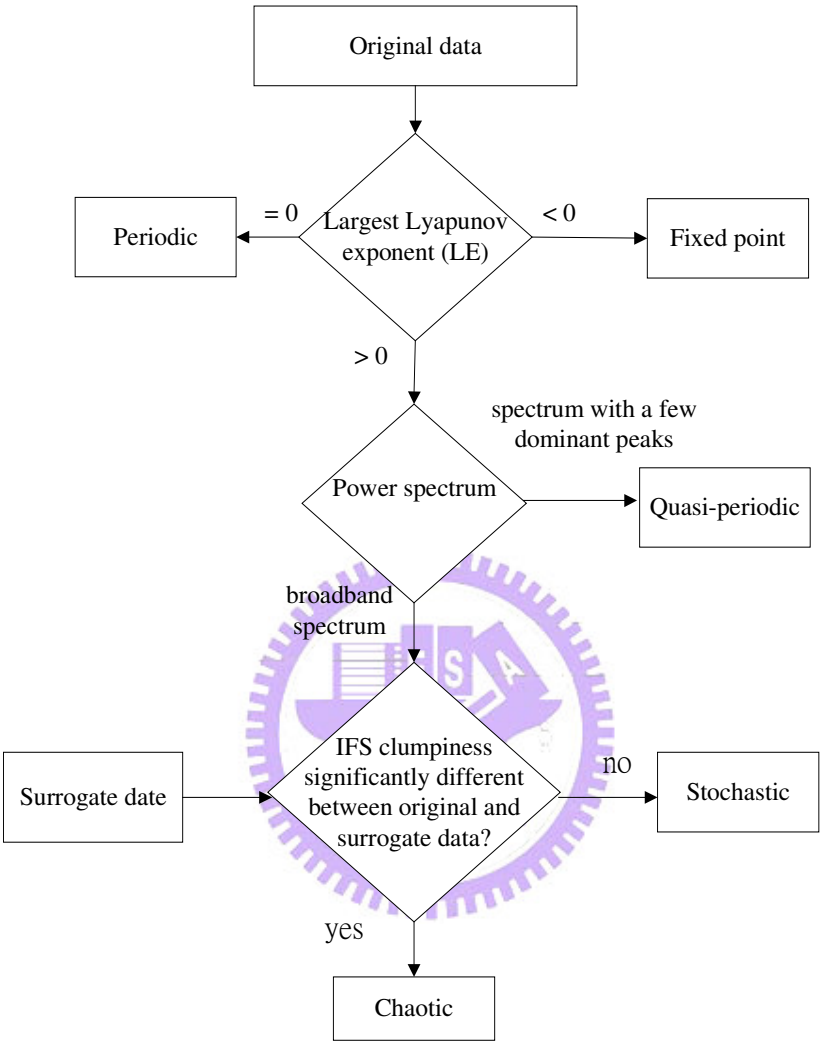
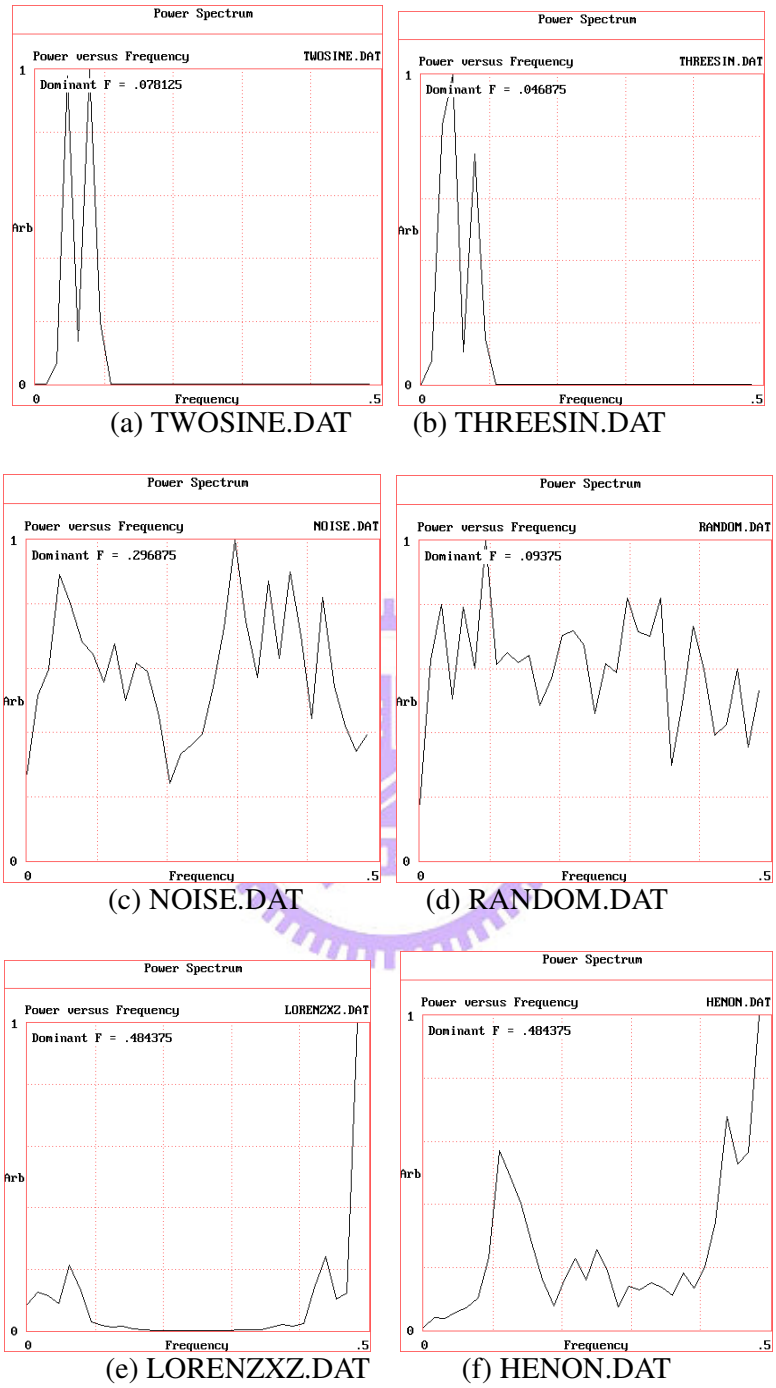



Figure 4-13 Proposed parsimony procedure

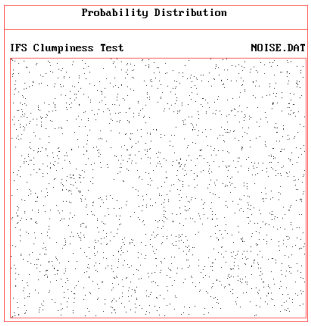


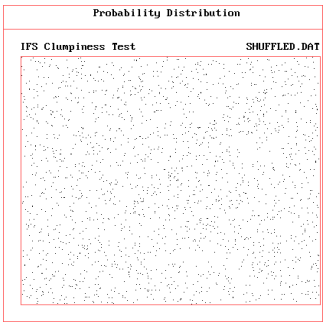

(a) TWOSINE.DAT (b) THREESIN.DAT

(c) NOISE.DAT (d) RANDOM.DAT

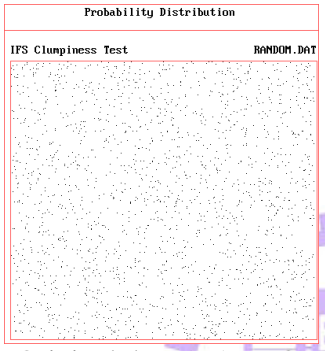
(e) LORENZXZ.DAT (f) HENON.DAT

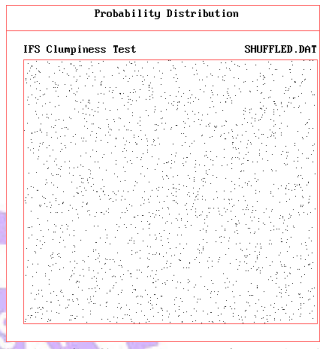
Figure 4-14 Power spectra for known time series generators



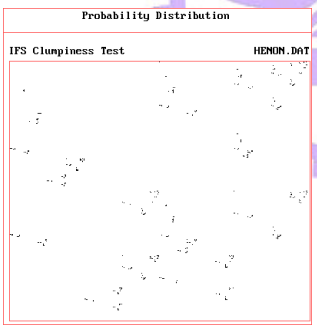



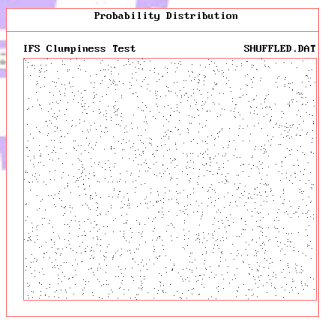

(a) Original data (NOISE.DAT) (b) Surrogate data (NOISE.DAT)



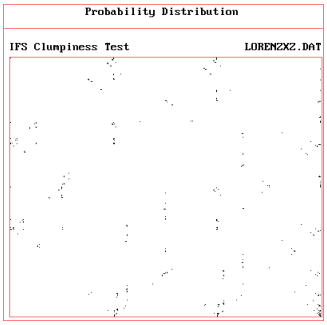



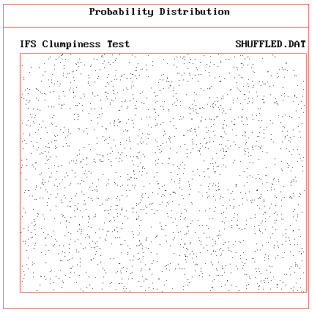

(b) Original data (RANDOM.DAT) (d) Surrogate data (RANDOM.DAT)




(e) Original data (HENON.DAT) (f) Surrogate data (HENON.DAT)




(g) Original data (LORENZXZ.DAT) (h) Surrogate data (LORENZXZ.DAT)

Figure 4-15 IFS clumpiness maps for known time series generators

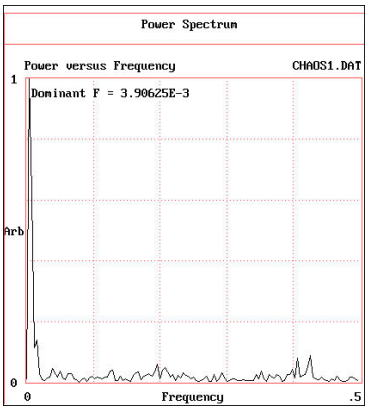
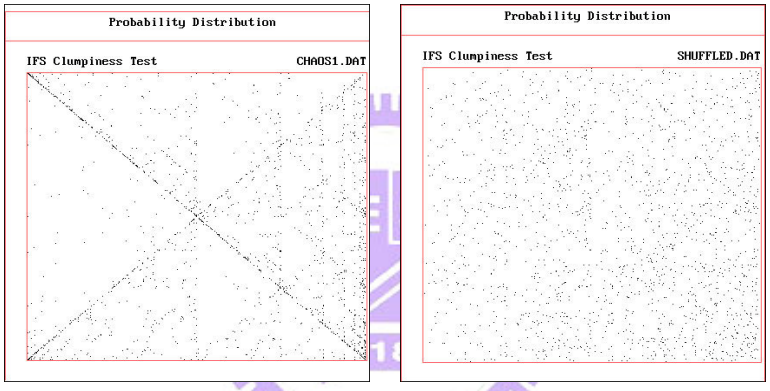



Figure 4-16 Power spectrum for station 32




(a) Original data (b) Surrogate data

Figure 4-17 IFS clumpiness maps for station 32