

# 连续动力系统时间序列的非线性检验<sup>\*</sup>

雷 敏<sup>1)</sup>† 孟 光<sup>1)</sup> 冯正进<sup>2)</sup>

<sup>1)</sup> 上海交通大学振动、冲击、噪声国家重点实验室, 上海 200030)

<sup>2)</sup> 上海交通大学机电控制研究所, 上海 200030)

(2004 年 6 月 21 日收到, 2004 年 7 月 15 日收到修改稿)

用相位随机化的替代数据方法, 研究了连续混沌动力系统时间序列的非线性检验判定. 研究发现, 在不同的采样情况下, 混沌时间序列的非线性特性检验有所差异, 尤其对于过采样时间序列, 往往会出现虚假的判断结果. 针对过采样时间序列, 最好采用能够反映非线性特征的参数, 作为检验统计量进行检验判断.

关键词: 替代数据, 关联维数, 连续时间序列, 非线性检验

PACC: 0540, 0545

## 1. 引 言

替代数据方法是一类用于时间序列非线性检验的方法, 也是一类间接判断原始信号混沌动力学特性的方法. 它自产生以来已被广泛地应用<sup>[1-6]</sup>和发展<sup>[7-11]</sup>. 这类方法由零假设和检验统计量两部分组成. 其基本思想是首先指定某种线性随机过程为零假设, 并依据该假设产生相应的一组替代数据, 然后分别计算原始数据和替代数据集的检验统计量, 最后根据原序列和替代数据统计量的显著性差异水平, 在一定置信度内决定接受或拒绝零假设<sup>[7, 12]</sup>. 目前, 基于 FT 算法的替代数据方法(又称为相位随机化的替代数据方法)因其简便可靠而被广泛地使用.

在现实生活中, 存在各种各样的时间序列问题, 它们往往是貌似随机信号的非线性(甚至混沌)数据. 因此, 判断时间序列的特性是进一步分析研究数据的前提. 但这些时间序列数据通常都来自相应的连续动力系统, 从而在判断其特性时应考虑采样间隔对判断结果的影响. 本文利用基于 FT 算法的替代数据方法, 对不同采样情况下的时间序列进行了检验研究.

## 2. 基于 FT 算法的替代数据方法

### 2.1. 零假设及其算法

零假设<sup>[7]</sup> 观测数据是由具有原始数据的均值和方差的线性相关高斯过程所产生. 若待测数据和替代数据的检验统计量之间差异显著, 该零假设不成立, 说明原始数据中必定包含非线性成分; 反之, 说明原始数据与替代数据一样, 都是由线性随机过程产生的.

依据这一零假设的替代数据, 可由两种方法产生, 一种方法是直接利用自回归 AR 模型产生:

$$x_t = a_0 + \sum_{k=1}^q a_k x_{t-k} + \sigma e_t, \quad (1)$$

即通过方程的不断迭代而得到替代数据, 但是用实验数据拟合(1)式中系数时, 会产生误差, 致使迭代结果很容易发散, 所以这种算法很不稳定; 另一种方法是 Theiler 等人<sup>[7]</sup>采用的将 Fourier 变换结果的相位进行随机化处理的方法, 这种算法较稳定. 根据 Wiener-Khinchin 定理, 这两种算法本质上等价. 这种方法的的思想是, 通过重构原始数据的功率谱, 以保证替代数据与原始数据的线性相关性, 故该相位随机化替代数据保持了原时间序列一样的功率谱、均值、方差和自相关函数等特征, 但具有不同的幅值直

<sup>\*</sup> 国家自然科学基金(批准号: 50335030)和国家高技术研究发展计划(批准号: 2002AA412410)资助的课题.

<sup>†</sup> E-mail: leimin@sjtu.edu.cn

方图.

产生替代数据的步骤<sup>[13]</sup>是先对实验数据进行 Fourier 变换,设观测数据为  $x(n)$ ,则它的离散 Fourier 变换为

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-2\pi i n k / N}, \quad (2)$$

接着,对所得各个频率处的变换值乘以  $e^{i\varphi}$ ,进行相位随机化处理:

$$X'(k) = X(k) e^{i\varphi}, \quad (3)$$

其中随机化相位区间选为  $[-\pi, \pi]$ ,并且相位的斜对称条件将根据实际数据个数的奇偶性而有所变化,当数据为偶数时,  $\phi(f_0) = 0, \phi(f_i) = -\phi(f_k) \chi i = 2 - N/2, k = N - N/2 + 1, \phi(f_{N/2+1}) = 0$ ;当数据为奇数时,  $\phi(f_0) = 0, \phi(f_i) = -\phi(f_k) \chi i = 2 - (N + 1)/2, k = N - (N + 1)/2 + 1$ .这样再将随机化相位后的 Fourier 谱数据进行 Fourier 逆变换:

$$x'(n) = \frac{1}{N} \sum_{k=0}^{N-1} X'(k) e^{2\pi i n k / N}, \quad (4)$$

$x'(n)$ 即为所求的替代数据.

## 2.2. 检验统计量

原则上,零假设是可以与任何检验统计量一起使用,即对于所有拒绝零假设的数据,其检验统计量分布与替代数据的不同.但一些文献指出并非所有检验统计量都一样有效,因此应该谨慎地选择有效的检验统计量<sup>[12]</sup>.

本文利用如下两种检验统计量对不同采样时间序列进行检验研究.一种是由(5)式给出的检验统计量  $T$ ,它不受均值和方差影响,可以判断除均值和方差外的一些线性结构特征是否一致.

$$T = \overline{(x - \bar{x})^4} / \overline{(x - \bar{x})^2}^2, \quad (5)$$

其中  $\bar{x}$  为  $x$  的均值.另一种是用关联维数作检验统计量  $T$ .关联维数是一种描述混沌吸引子分形结构的量化指标,测量了系统的自由度和复杂度,是一种反映非线性特征的参数.设点  $X_1, X_2, \dots, X_n$  为相空间内吸引子上的点,则关联维数<sup>[14]</sup>为

$$D_2 = \lim_{l \rightarrow 0} \frac{\ln \alpha(l)}{\ln(l)}, \quad (6)$$

$$\alpha(l) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \theta(l - \|X_i - X_j\|), \quad (7)$$

其中  $\|\cdot\|$  为 Euclidean 范数,  $\theta$  为 Heaviside 函数.相应的检验统计量为

$$T = D_2. \quad (8)$$

## 3. 不同采样的时间序列检验

本文以 Lorenz 混沌连续系统(9)为例<sup>[15]</sup>,利用替代数据方法,研究了不同采样情况的时间序列特性.

$$\begin{cases} \frac{dx}{dt} = \sigma(y - x), \\ \frac{dy}{dt} = rx - y - xz, \\ \frac{dz}{dt} = -bz + xy, \end{cases} \quad (9)$$

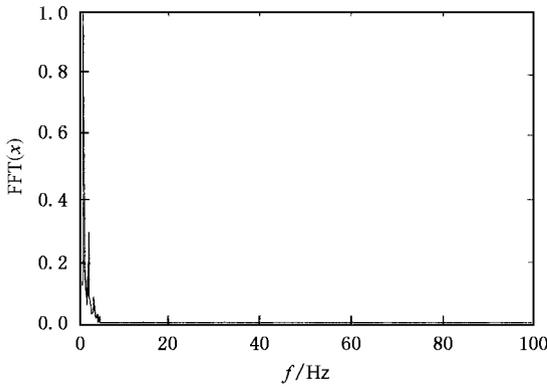
其中 Lorenz 系统的参数为  $\sigma = 10, b = 8/3, r = 28$ ;初始状态为  $x(0) = 5, y(0) = 5, z(0) = 15$ .混沌时间序列为状态变量  $x$  的采样值,采样点数为 1000 点.为了使零假设在 95% 置信度范围内有效,应产生 39 组替代数据(即若置信度为  $p$ ,则替代数据集  $B$  最小为<sup>[12]</sup>  $B_{\min} = 2(1 - p) - 1$ ,故当  $p = 95\%$  时,  $B_{\min} = 39$ ),这里检验统计量由(5)式给出.

### 3.1. 过采样情况

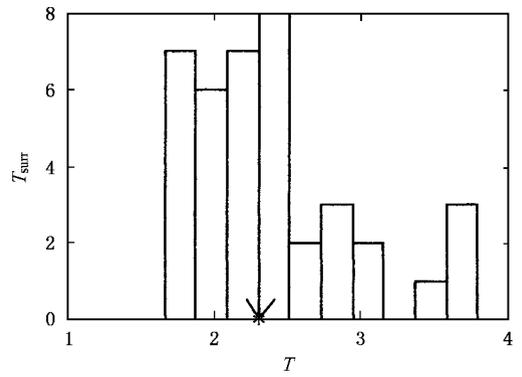
根据 Shannon 采样定理可知,当采样频率远远大于信号的两倍主要最大频率  $f_{s \max}$  时,此时所采集的时间序列为过采样情况下的数据.图 1(a)为采样间隔  $\tau = 0.005$  s(采样频率  $f_s = 200$  Hz)时的 Lorenz 混沌时间序列  $x$  频谱图.从图 1 可以看出,采样频率 ( $f_s = 200$  Hz)远远大于 Lorenz 混沌时间序列  $x$  的主要最大频率( $f_{s \max}$  约为 5 Hz),故当采样间隔  $\tau = 0.005$  s 时, Lorenz 混沌时间序列  $x$  为过采样数据.图 1(b)为相应的过采样 Lorenz 混沌时间序列  $x$  的替代数据检验结果.可以看出,在过采样情况下,原始数据的检验统计量  $T_{\text{orig}}$  与替代数据的检验统计量  $T_{\text{sur}}$  无明显差异.根据替代数据的零假设,在 95% 的置信度范围内,该零假设不能被拒绝,原始混沌时间序列  $x$  被判断成线性的,这显然是虚假的结果.接下来,本文又对欠采样情况和合适采样情况进行了研究.

### 3.2. 欠采样情况

根据 Shannon 采样定理,当采样频率小于信号的两倍主要最大频率时,为欠采样情况.从图 2(a)可以看出,当采样间隔  $\tau = 1$  s(采样频率  $f_s = 1$  Hz)时,采样频率  $f_s$  小于 Lorenz 混沌时间序列  $x$  的两倍主要最大频率  $f_{s \max}$ ,故此时的 Lorenz 混沌时间序列



(a)  $\tau = 0.005s$   $f_s = 200Hz$

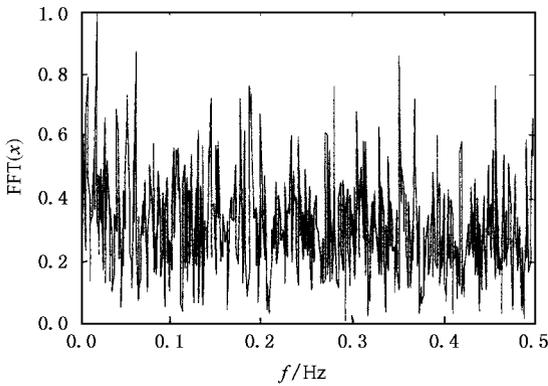


(b) \* 为  $T_{orig}$  值, 直方图为  $T_{surr}$  值

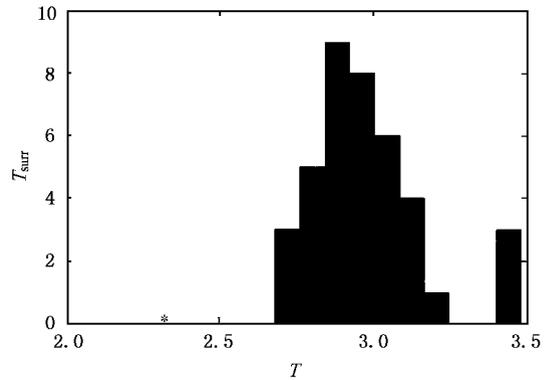
图 1 过采样情况下, Lorenz 混沌时间序列  $x$  的分析和检验结果

$x$  为欠采样数据. 在这种情况下, 由图 2(b) 可以看出, Lorenz 混沌时间序列  $x$  的检验统计量  $T_{orig}$  与相应的替代数据的检验统计量  $T_{surr}$  之间有显著差异, 根

据替代数据的零假设, 该零假设在 95% 的置信度范围内被拒绝, 故可以判断 Lorenz 混沌时间序列  $x$  存在非线性成分.

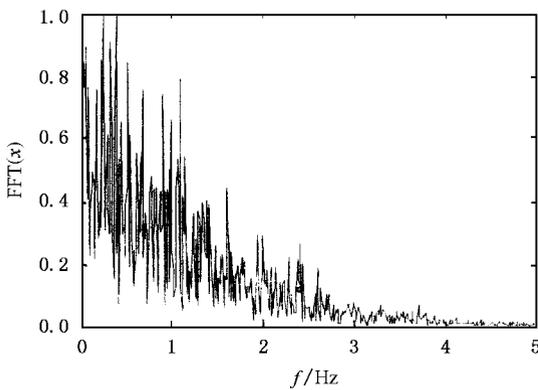


(a)  $\tau = 1s$   $f_s = 1Hz$

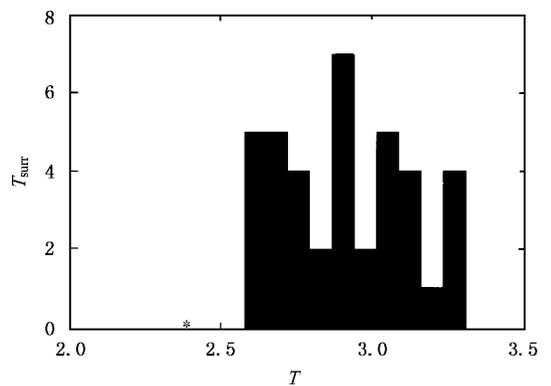


(b) \* 为  $T_{orig}$  值, 直方图为  $T_{surr}$  值

图 2 欠采样情况下, Lorenz 混沌时间序列  $x$  的分析和检验结果



(a)  $\tau = 0.1s$   $f_s = 10Hz$



(b) \* 为  $T_{orig}$  值, 直方图为  $T_{surr}$  值

图 3 合适采样情况下, Lorenz 混沌时间序列  $x$  的分析和检验结果

### 3.3. 合适采样情况

对于 Lorenz 混沌时间序列  $x$  而言,除了上述采样情况外,其采样数据为合适采样,如图 3(a)所示(采样间隔  $\tau = 0.1s$ ,  $f_s = 10Hz$ ).在这种情况下,从图 3(b)可以看出,Lorenz 混沌时间序列  $x$  的检验统计量  $T_{orig}$  与替代数据的检验统计量  $T_{sur}$  存在显著差异,根据替代数据的零假设,该零假设在 95% 的置信度范围内被拒绝,可以判断 Lorenz 混沌时间序列  $x$  存在非线性成分.

## 4. 讨论与结论

从上一节的研究结果可以看出,用(5)式给出的检验统计量对连续混沌动力系统的时序进行非线性检验时,有时会给出虚假的判断结果.即在过采样情况下,基于 FT 算法的替代数据方法,不能有效地检验出混沌时间序列的非线性特性.这主要是当数据为过采样情况时,数据中的线性相关成分远远大于非线性成分,因此其特性主要体现为线性的(尤其对于有限长度的混沌时间序列),并可以用线性模型(如 AR 模型)进行描述(如图 4 所示),而(5)式主要反映数据中线性特性的一致性,从而很难用该算式检验出过采样时间序列的非线性特性.这一现象,在其他连续混沌动力系统(如 Rossler 混沌系统、Duffier 混沌系统<sup>[15]</sup>,甚至文献[16]中所提及的  $M =$

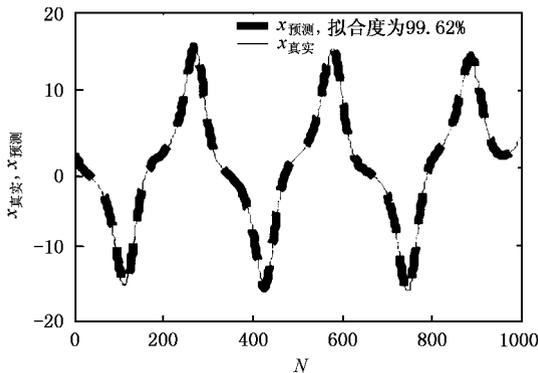


图 4  $\tau = 0.005s$  时,Lorenz 混沌时间序列  $x$  及其 10 阶 AR 模型预测结果

101 超混沌系统)中也同样存在.因此在过采样情况下,最好选用能够反映非线性特性的参数,作为检验统计量进行判断,如分形维数、Lyapunov 指数、熵、复杂度等.本文选用关联维数作为检验统计量,对上述过采样 Lorenz 混沌时间序列进行了重新检验研究.从图 5 可以看出,原始数据与替代数据之间存在显著差异,零假设被拒绝,原始数据存在非线性成分,从而得到了很好的检验结果,也进一步验证了该思路的正确性.

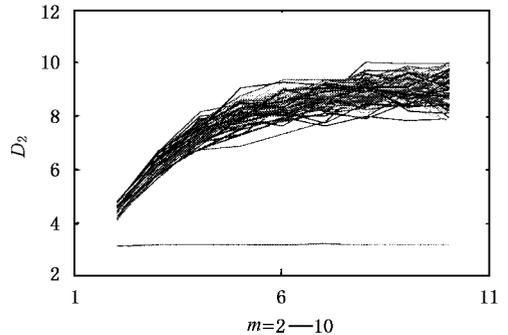


图 5 检验统计量为关联维数时,过采样间隔情况下 Lorenz 混沌时间序列的检验分析

当然,对于欠采样和合适采样情况,关联维数为检验统计量时,同样也能检验出混沌时间序列的非线性特性.但需要指出的是,这种检验方法非常费时,如果用其他检验方法可以有效检验,就不必用这种检验方法.另外,本文为了快速获取各种情况下的关联维数,将所有关联积分曲线中的直线部分取为一样.这样,不仅能有效地检验混沌时间序列,也避免了人为因素的影响.

综上所述,对于连续系统时间序列,如果可以用简单类型的检验统计量(如(5)式)有效地检验其非线性特性,就不必使用非常费时的复杂类型的检验统计量(如关联维数)进行检验.这样,可减少检验所需的时间,从而可为需要实时检验判断的场合提供方便可靠的非线性检验手段.对于过采样时间序列,最好利用复杂类型的检验统计量进行有效地检验.本文的研究为分析实际连续动力系统时间序列的非线性特性提供了理论参考.

[1] Poon C S and Merrill C K 1997 *Nature* **389** 492

[2] Parlitz U and Kocarev L 1997 *Int. J. Bifurc. Chaos* **7** 407

- [ 3 ] Small M and Judd K 1998 *Int. J. Bifurc. Chaos* **8** 1231 等 2001 物理学报 **50** 633 ]
- [ 4 ] Zhang H X *et al* 2000 *Acta Phys. Sin.* **49** 1416( in Chinese ) [ 张红 焯等 2000 物理学报 **49** 1416 ]
- [ 5 ] Xie Y *et al* 2002 *Acta Phys. Sin.* **51** 205( in Chinese ) [ 谢 勇等 2002 物理学报 **51** 205 ]
- [ 6 ] Xie Y *et al* 2003 *Acta Phys. Sin.* **52** 1112( in Chinese ) [ 谢 勇等 2003 物理学报 **52** 1112 ]
- [ 7 ] Theiler J *et al* 1992 *Physica D* **58** 77
- [ 8 ] Schreiber T and Schmitz A 1996 *Phys. Rev. Lett.* **77** 635
- [ 9 ] Prichard D and Theiler J 1994 *Phys. Rev. Lett.* **73** 951
- [ 10 ] Liu Y Z *et al* 2001 *Acta Phys. Sin.* **50** 633( in Chinese ) [ 刘耀宗 等 2001 物理学报 **50** 633 ]
- [ 11 ] Liu Y Z *et al* 2001 *Acta Phys. Sin.* **50** 1241( in Chinese ) [ 刘耀宗 等 2001 物理学报 **50** 1241 ]
- [ 12 ] Theiler J and Prichard D 1996 *Physica D* **94** 221
- [ 13 ] Lei M , Wang Z Z and Feng Z J 2001 *Phys. Lett. A* **290** 297
- [ 14 ] Grassberger P and Procaccia I 1983 *Phys. Rev. Lett.* **50** 346
- [ 15 ] Wang D S *et al* 1995 *Chaos , Fractal and Its Applications*( Hefei : Pross of University of Science and Technology of China ) p157 , 176 , 181( in Chinese ) [ 王东生等 1995 混沌、分形及其应用(合肥 : 中国科学技术大学出版社 )第 157 , 176 , 181 页 ]
- [ 16 ] Baier G and Sahle S 1995 *Phys. Rev. E* **51** R2712

## Detecting the nonlinearity for time series sampled from continuous dynamic systems<sup>\*</sup>

Lei Min<sup>1)†</sup> Meng Guang<sup>1)</sup> Feng Zheng-Jin<sup>2)</sup>

<sup>1)</sup> *State Key Laboratory of Vibration , Shock & Noise , Shanghai Jiaotong University , Shanghai 200030 , China )*

<sup>2)</sup> *Institute of Mechatronic Control System , Shanghai Jiaotong University , Shanghai 200030 , China )*

( Received 21 June 2004 ; revised manuscript received 15 July 2004 )

### Abstract

This paper studies the detection of the nonlinearity of time series sampled from continuous dynamics systems by using the surrogate data method. The results show that under the different sampling conditions , the detection finds different nonlinearity of chaotic time series. Especially for the oversampling time series , there can often be some illusive results. For this , we suggest that it is best to apply nonlinear values as testing statistics for detecting nonlinearity of oversampling time series.

**Keywords** : surrogate data , correlation dimension , nonlinear time series , nonlinear testing

**PACC** : 0540 , 0545

<sup>\*</sup> Project supported by the National Natural Science Foundation of China ( Grant No. 50335030 ) , and the National High Technology Development Program of China ( Grant No. 2002AA412410 ).

<sup>†</sup> E-mail : leimin@sjtu.edu.cn