

一种多变量时间序列的分形维数计算方法

吴虎胜¹, 倪丽萍², 张凤鸣¹, 周 漩¹, 杜继永¹

(1. 空军工程大学 装备管理与安全工程学院, 西安 710051; 2. 合肥工业大学 管理学院, 合肥 230009)

摘 要: 分形维数是描述混沌动力学系统的重要参数之一. 根据时间尺度与多维超体体积之间的测度关系, 提出一种多变量时间序列分形维数的计算方法. 通过4种典型混沌动力学系统所产生的多变量时间序列及其相应不同信噪比混杂序列的仿真计算表明, 所提出方法时间复杂度较低, 所需序列长度较短, 具有一定的抗噪能力, 且无需进行相空间重构, 避免了嵌入维数和延迟时间等参数选取对结果造成的影响, 是计算多变量时间序列分形维数的一种有效途径.

关键词: 多变量时间序列; 分形维数; 混沌系统

中图分类号: TP13; O415

文献标志码: A

A method for calculating fractal dimension of multivariate time series

WU Hu-sheng¹, NI Li-ping², ZHANG Feng-ming¹, ZHOU Xuan¹, DU Ji-yong¹

(1. Materiel Management and Safety Engineering College, Air Force Engineering University, Xi'an 710051, China;

2. School of Management, Hefei University of Technology, Hefei 230009, China. Correspondent: WU Hu-sheng,

E-mail: wuhusheng0421@163.com)

Abstract: The fractal dimension is an important parameter of detecting and characterizing chaos produced from a dynamical system. According to the relationship between time scale and multidimensional super-body's volume, a method to compute the fractal dimension of multivariate time series is proposed. Numerical simulations show that the length of the time series is shorter, the method is more efficient and doesn't need reconstructing phase space so as to avoid the infection from parameters determination such as delay time and embedding dimension. Furthermore, it can resist noise to a certain degree and can be used as an effective method to calculate the fractal dimension of multivariate time series in many areas.

Key words: multivariate time series; fractal dimension; chaotic system

0 引 言

现实中, 多变量时间序列广泛存在于医学、金融、国防等各领域, 由其所刻画的很多系统也都具有分形特征^[1-2]. 分形维数是描述混沌动力学系统的重要参数, 能促进研究者对事物物理本质的认识. 目前的研究主要集中于单变量时间序列中, 而从理论上讲, 多变量时间序列包含原系统更加丰富的信息, 采用其计算所得的分形维数应更加真实^[3]. 当前, 计算多变量时间序列分形维数主要有基于相空间重构的方法和原形法. 对于前者, 由 Takens 定理, 只要数据量足够、参数选取合适, 混沌系统中的单变量时间序列即可重构原系统, 进而可在高维相空间中计算分形维数. 但事实并非完全如此, 例如, 若令 Lorenz 系统方程中的 $x = -x, y = -y, z = z$, 则方程仍成立, 那么 X

轴与 Y 轴的对称关系就没能反应到 Z 轴数据中, 即由 z 坐标不能分辨 x, y 的对称性, 故由 z 坐标的测量不能重构该系统, 从而不能保证实际问题中任何单变量时间序列都可重构原系统^[4]. 同时, 现实中也很难保证待计算的多变量时间序列属于同一混沌动力学系统.

鉴于此, 学者们研究了基于多变量时间序列相空间重构的方法, 其基本思想是首先对每个单变量时间序列分别确定最佳延迟时间和嵌入维数以重构相空间, 然后通过广义关联积分求得分形维数^[5]. 相对于基于单变量时间序列相空间重构的方法, 所需时间序列的长度短且准确度高^[6], 但也大大增加了计算量. 同时引入了较为复杂的多变量时间序列相空间重构的参数选择问题, 导致即使对于同组数据, 由于参数

收稿日期: 2012-11-08; 修回日期: 2013-04-01.

基金项目: 国家自然科学基金项目(60304004).

作者简介: 吴虎胜(1986-), 男, 博士生, 从事信息系统工程与智能决策、智能数据挖掘的研究; 倪丽萍(1981-), 女, 副教授, 博士, 从事数据挖掘、人工智能等研究.

的选择不同而使得计算结果存在较大差异. 原形法则无需进行相空间重构, 它将多变量时间序列看作一个多元数据集直接求取其维数. 该方法适用范围较广, 文献[7]指出此类方法甚至可用于分析某些不是严格分形意义上的数据. 其典型代表是盒计数法, 计算简单、易于理解, 但其也存在空盒计数问题^[8], 影响分形维数的计算精度.

本文根据多变量时间序列的时间尺度与其超体体积的关系绘制双对数曲线, 并结合一种分段拟合误差法自动识别无标度区, 进而给出一种无需进行相空间重构的多变量时间序列分形维数计算方法, 并分析了算法的时间复杂度和噪声对结果的影响. 通过 Lorenz 系统、Chen's 系统、Rossler 系统和 4 阶超混沌系统的仿真计算, 验证了所提出算法的有效性.

1 分形维数的计算

1.1 方法思想

若将一立方体的边长 L 都扩大到原来的 k 倍, 则其二维测度表面积 S 为原来的 k^2 倍, 三维测度的体积 V 为原来的 k^3 倍, 即存在以下关系式:

$$L \propto S^{1/2} \propto V^{1/3}. \quad (1)$$

若假设 Y 为具有 D 维测度的量, 则可将式 (1) 变为

$$L \propto S^{1/2} \propto V^{1/3} \propto Y^{1/D}. \quad (2)$$

因此, 可以根据上述测度关系求取分形维数^[9]. 以经典的英国海岸线问题为例, 设面积为 S , 海岸线长度为 X_{line} , S 是具有二维测度的量, 由 $S^{1/2} \propto X_{\text{line}}^{1/D}$ 即可求出海岸线的分形维数 D , 同样也可用海岸线长度 X_{line} 与海岸线两端直线距离 L 之间的关系计算维数 D .

分形理论初期应用于描述自然界和非线性系统中的不光滑或不规则的几何形体, 这些几何形体在某种意义上具有图形或结构上的自相似性. 对于时间序列, 这种自相似性主要体现在时间上. 例如, 分析十余年的上证指数的日收盘价曲线, 可以发现其随时间的变化在形态上呈现出一定的相似性^[10]. 对于多变量时间序列, 在某一时间段, 序列中的点所占据的空间可以用多维超体测度^[11]. 具体计算时, 将时间跨度作为测量尺度, 若多变量时间序列具有混沌特性, 则时间尺度 r 与多维超体的体积 V 之间应该呈现某种意义上的自相似性, 进而可以依据尺度与超体体积的测度关系 $r \propto V^{1/D}$ 求取分形维数.

1.2 超体体积计算

如上述分析, 假设多变量时间序列每一变量维度的测量尺度扩大至原来的 r 倍, 则相应的多维超体体积尺度就扩大 r^m 倍. 当 $m > 1$ 时, 按照时间先后顺序记录的值 $TS_i = \{x_i(0), x_i(1), \dots, x_i(t), \dots, x_i(n)\}$

称为多变量时间序列. 其中: t 为时间点, n 为时间序列长度, i 为变量, $x_i(t)$ 为第 i 个变量在 t 时刻上的记录值. 多变量时间序列在时间尺度 r 下的超体体积为

$$V(r) = \frac{\sum_{q=0}^p \prod_{i=1}^m \left\{ \max_{rq \leq t \leq r(q+1)} [x_i(t)] - \min_{rq \leq t \leq r(q+1)} [x_i(t)] \right\}}{r^m}. \quad (3)$$

其中: m 为变量数; P 为时间尺度将序列分成了 $P+1$ 个部分, $P = \lceil (n-1)/r \rceil - 1$, $\lceil \cdot \rceil$ 表示取整;

$$\max_{rq \leq t \leq r(q+1)} [x_i(t)] - \min_{rq \leq t \leq r(q+1)} [x_i(t)]$$

为在尺度 r 下第 q 部分多维超体的第 i 维边长, 由第 i 维变量的子序列 $x_i(rq), x_i(rq+1), \dots, x_i(r(q+1))$ 中的最大值与最小值之差决定. 最后, 不断变换尺度 r 即可得到一系列 $V(r)$.

1.3 无标度区的确定

具有分形特征的数据应服从幂律分布 $V(r) = Cr^{\pm D}$ (C 为常数), 于是可拟合 $\log(r) \sim \log(V(r))$ 双对数曲线上的点, 得到的直线斜率绝对值即为所求的分形维数. 现实中, 大多系统并非存在严格意义上的分形特性, 而是近似或统计意义上存在, 且仅存在于一定范围内, 即所谓的无标度区. 如何确定无标度区是研究的关键, 这里设计一种分段拟合误差法自动识别无标度区, 即通过长度可变的窗口, 从序列的初始点开始, 窗口长度每增长一个单位就对窗内的点进行最小二乘拟合并计算拟合误差, 直至拟合误差大于某一给定的最大误差时, 记录此窗口内的序列为一分段; 然后在下一个点开始生成新窗口, 重复以上过程直至序列完毕; 最后统计各分段包含的点数, 点数最多的即为无标度区.

计算过程如下: 将得到的一系列尺度 r 和超体体积 $V(r)$ 分别取对数, 并令 $\ln r_j = R_j$, $\ln[V(r_j)] = V_j$, $j = 1, 2, \dots, Q$, 则双对数曲线上的点序列可表示为 $(R_1, V_1), (R_2, V_2), \dots, (R_Q, V_Q)$. 设某窗口的对应区间为 $[S_k, f_k]$, 分别表示第 k 个窗口在原序列中的起点和终点位置. 其中: $S_1 = 1, f_1 = 2, f_a = Q, a$ 为分段数. 假设 (R_j, V_j) 和 (R_j, V_j^*) 分别表示窗口内原子序列和其对应拟合直线序列上的第 j 个点, $j = S_k, S_k + 1, \dots, f_k$, 则此窗口内的拟合误差可定义为

$$e_k = \left[\sum_{j=S_k}^{f_k} (V_j - V_j^*)^2 \right]^{\frac{1}{2}}. \quad (4)$$

比较 e_k 与最大允许误差 e_{\max} : 当 $e_k < e_{\max}$ 时, $S_k = S_k, f_k = f_k + 1, k = k$, 拓展原窗口的长度, 再进行拟合误差计算; 当 $e_k \geq e_{\max}$ 时, $S_{f_k} = f_k - S_k + 1, k = k + 1, S_k = f_{k-1}, f_k = f_{k-1} + 1$ (S_{f_k} 为序列第 k 个分段包含的点数), 视此窗口内的子序列为一个分

段,另生成新窗口重新开始拟合计算直至到序列完毕. 参数 e_{\max} 可利用以下估计公式计算得到:

$$e_{\max} = \sqrt{\frac{1}{Q-1} \sum_{j=1}^Q (V_j - \bar{V})^2 / \bar{V}}, \quad \bar{V} = \sum_{j=1}^Q V_j / Q. \tag{5}$$

最后,统计各分段所包含点数,选取点数最多的分段进行直线拟合,其斜率的绝对值即为所求的多变量时间序列的分形维数.

1.4 算法复杂度分析

算法的时间复杂度主要分为两个部分:超体体积计算和确定无标度区.在超体体积计算中,对于每个尺度 r_j ,计算该尺度下多变量时间序列的超体体积 $V(r_j)$.

首先,尺度 r_j 将多变量时间序列分成 $[(n-1)/r_j]$ 个部分.对于每部分,需要对每个变量维度求取其在该尺度下的边长,即该尺度下每个变量维度的最大值与最小值之差,共计算 $2(r_j-1) \times m$ 次.因此,多变量时间序列的超体体积的复杂度为 $o\left(\sum_{j=1}^Q 2(r_j-1) \times m \times [(n-1)/r_j]\right)$.在确定无标度区时,计算 e_{\max} 的时间复杂度为 $o(2Q+1)$,确定此拟合直线的最差时间复杂度为 $o(Q \times (Q+1)/2)$.由于一般情况下 Q 远小于 n ,本算法的时间复杂度为 $o(Q \times m \times n)$.经典的基于相空间重构法计算分形维数的算法复杂度为 $o(K \times M \times n^2)$ [12].其中: K 为半径 R 取值的个数, M 为嵌入维数, n 为时间序列长度.对于原形法,以经典的盒计数法为例,计算复杂度为 $o(m \times n \times \log(m \times n))$ [13-14].可见,本文算法的复杂度相对较低.

2 仿真分析

2.1 算法有效性分析

为了表明本文方法的有效性,对 Lorenz 系统、Chen's 系统、Rossler 系统和一种 4 阶超混沌系统所产生的多变量时间序列进行仿真分析.将得到的结果与文献 [15-18] 方法从分形维数数值、所需时间序列长度等方面进行对比.实验所用的混沌动力学系统如表 1 所示,采用 4 阶 Runge-Kutta 法求解微分方程组以生成多变量时间序列.表 1 中, D 为本文方法的分形维数计算结果, $D^{[*]}$ 为文献 [*] 中相应混沌系统的分形维数计算结果, h 为步长, Q 为最大时间尺度, Len 为时间序列长度, Num 为无标度区所包含的序列点数, Fin 为无标度区拟合直线的拟合优度,参数 e_{\max} 皆由式 (5) 计算得到.

如表 1 所示,本文计算得到的分形维数与其他文献采用基于相空间重构方法或盒计数法所求的维数在数值上非常接近.由于参数 e_{\max} 可以由公式估计得到,且无需进行相空间重构,不仅减少了计算复杂度,而且避免了嵌入维数、延迟时间等参数选取对结果产生的影响,提高了结果的准确性.另外,本文方法由相对较短的时间序列即可精确地计算出分形维数,如,计算 Lorenz 系统产生多变量时间序列的分形维数时,文献 [15] 所用时间序列长度为 50 000,本文方法所用序列长度为 2 000,相比而言,本文所用序列的长度要短且计算结果也很接近.

2.2 算法抗噪性能分析

以 Rossler 系统(图 1(a)所示)为例,将其产生的多变量时间序列混入不同信噪比的高斯白噪声后

表 1 数值实验所用的混沌动力学系统及其计算结果对比

名称	方程	参数值	初值	h	e_{\max} 和 Q	Fin 和 Num	D 和 Len	$D^{[*]}$ 和 Len
Lorenz	$\dot{x} = \sigma(y - x)$	$\sigma = 10$	$x_0 = 15.34$	0.01	0.42	99.38%	1.971 5	2.040 9 ^[15]
	$\dot{y} = \beta x - y - xz$	$\beta = 28$	$y_0 = 13.68$					
	$\dot{z} = xy - bz$	$b = 8/3$	$z_0 = 37.91$					
Chen's	$\dot{x} = a(y - x)$	$a = 35$	$x_0 = 0$	0.01	0.39	99.63%	2.363 0	2.367 6 ^[16]
	$\dot{y} = (c - a)x - xz + cy$	$b = 3$	$y_0 = 1.001$					
	$\dot{z} = xy - bz$	$c = 28$	$z_0 = 0$					
Rossler	$\dot{x} = -y - z$	$a = 0.2$	$x_0 = 1$	0.01	0.26	99.69%	1.782 4	1.826 3 ^[17]
	$\dot{y} = ay + x$	$b = 0.4$	$y_0 = 0$					
	$\dot{z} = b + z(x - c)$	$c = 5.7$	$z_0 = 0$					
4 阶超混沌		$a = 45$		0.001	0.28	99.94%	3.207 7	3.224 2 ^[18]
		$b = 16$						
	$\dot{x} = a(y - x) + eyz - ku$	$c = 48$	$x_0 = 0$					
	$\dot{y} = cx - dy - xz$	$d = 6$	$y_0 = 0$					
	$\dot{z} = xy - bz$	$e = 30$	$z_0 = 0$					
	$\dot{u} = rx + fyz$	$f = 35$	$u_0 = 0$					
		$k = 20$						
		$r = 35$						

表 2 噪声对各系统的分形维数计算的影响

混沌系统 信噪比	Lorenz				Chen's				Rossler			
	e_{\max}	Num	D	$E/\%$	e_{\max}	Num	D	$E/\%$	e_{\max}	Num	D	$E/\%$
30	0.42	35	1.971 8	3.39	0.39	49	2.355 9	0.49	0.29	64	1.795 1	1.71
20	0.42	35	1.974 8	3.24	0.46	55	2.477 7	4.65	0.41	71	1.922 0	5.24
10	0.42	32	2.035 1	0.28	0.45	57	2.512 1	6.10	0.91	98	2.319 5	27.01

得到混杂序列(见图 1(b)~图 1(d)). 可以看出, Rossler 系统加入噪声后的混杂序列仍可视作奇异吸引子, 但相对于原 Rossler 系统的吸引子, 由于噪声的影响, 其分形特征受到影响, 吸引子发生了不同程度的扭曲变形. 以 Lorenz, Chen's, Rossler 三个典型的混沌动力学系统为例进行说明, 所用系统的参数、初值设置同表 1. 将不同信噪比的混杂序列按本文方法计算其分形维数. 表 2 中, E 为混杂序列的分形维数与表 1 中

相应维数参考值的相对误差. 由表 2 可见, 随着信噪比的降低, 噪声功率的增大, 所计算的分形维数有逐渐增加的趋势. 这主要是由于加入的 3 维高斯白噪声序列的分形维数约为 2.7, 大于 3 种混沌动力学系统所产生多变量时间序列的分形维数. 因此, 噪声的影响使得信噪比低的混杂序列的分形维数偏大. 由于 3 种系统本身混沌特性的稳定性不一, 受噪声的影响程度不同, 噪声对其分形维数计算结果的影响也不同, 但除信噪比为 10 dB 的 Rossler 混杂多变量时间序列外, 其他多变量时间序列分形维数的计算结果均保持在较小的误差范围内, 这表明本文方法具有一定的抗噪能力和计算鲁棒性.

3 结 论

本文提出了一种多变量时间序列分形维数的计算方法. 仿真计算表明: 采用多变量时间序列计算分形维数的结果与单变量计算的结果基本一致, 但所用的数据量明显减少; 相对于多变量时间序列相空间重构后再计算分形维数的方法, 本文方法无需进行相空间重构, 因此也不用进行嵌入维数、延迟时间等参数的选取, 且算法时间复杂度相对较低并具有一定的抗噪能力. 该方法可应用到实际中, 如计算某证券系统的分形维数以定量刻画该系统的复杂程度; 计算脑电图等多变量时间序列的分形维数以区分正常状态和不同的病理状态, 进行人体系统的疾病诊断等.

参考文献(References)

- [1] Korn H, Faure P. Is there chaos in the brain? Experimental evidence and related models[J]. Computers Rendus Biologies, 2003, 326(9): 787-840.
- [2] Muge Iseri, Hikmet Caglar, Nazan Caglar. A model proposal for the chaotic structure of Istanbul stock exchange[J]. Chaos, Solitons and Fractals, 2008, 36(5): 1392-1398.
- [3] 占山, 王海燕. 多变量时间序列最大李雅普诺夫指数的计算[J]. 物理学报, 2006, 55(2): 572-575.
(Zhan S, Wang H Y. Calculation of the maximal Lyapunov exponent from multivariate data[J]. Acta Physica Sinica, 2006, 55(2): 572-575.)
- [4] Cao Liangyue, Mees A, Judd K. Dynamics from multivariate time series[J]. Physica D, 1998, 121(2): 75-88.

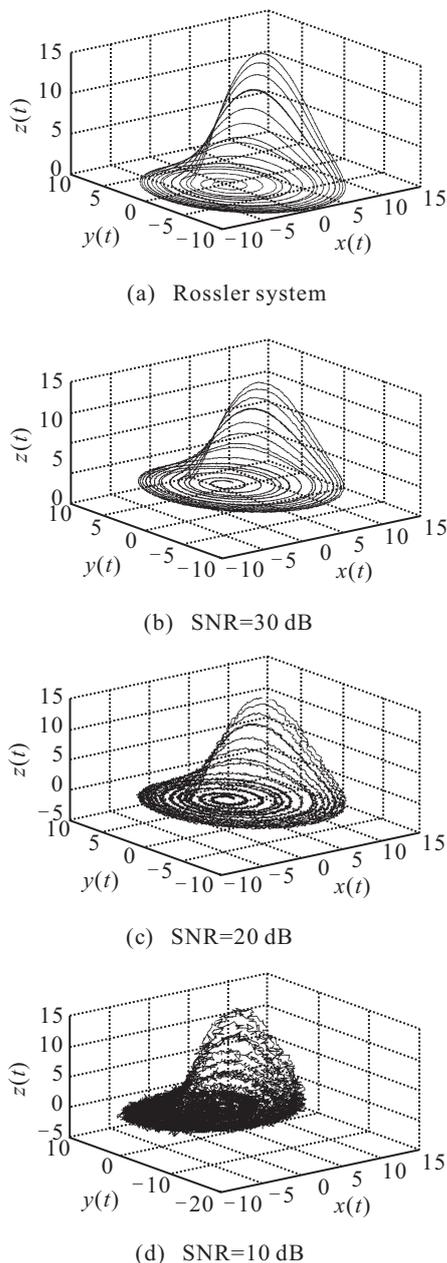


图 1 不同信噪比的混合多变量时间序列的吸引子

- [5] 岳毅宏, 韩文秀. 多变量时间序列相空间重构中参数的确定[J]. 控制与决策, 2005, 20(3): 290-293.
(Qiu Y H, Han W X. Determination of parameters in the phase-space reconstruction of multivariate time series[J]. Control and Decision, 2005, 20(3): 290-293.)
- [6] 马春涛, 马千里, 彭宏, 等. 基于条件熵扩维的多变量混沌时间序列相空间重构[J]. 物理学报, 2011, 60(2): 81-88.
(Ma C T, Ma Q L, Peng H, et al. Multivariate chaotic time series phase space reconstruction based on extending dimension by conditional entropy[J]. Acta Physica Sinica, 2011, 60(2): 81-88.)
- [7] Nikoli D, Moca V V, Singer W, et al. Properties of multivariate data investigated by fractal dimensionality[J]. J of Neuroscience Methods, 2008, 172(1): 27-33.
- [8] James Theiler. Estimating fractal dimension[J]. J of the Optical Society of America A, 1990, 7(6): 1055-1072.
- [9] 张济忠. 分形[M]. 北京: 清华大学出版社, 2011: 73-75.
(Zhang J Z. The fractal[M]. Beijing: Tsinghua University Press, 2011: 73-75.)
- [10] 姜灵敏, 周峰. 上证指数盒维数的计量与特性研究[J]. 系统工程学报, 2006, 21(4): 434-437.
(Jiang L M, Zhou F. Box-dimension measure and characteristics of Shanghai general index[J]. J of Systems Engineering, 2006, 21(4): 434-437.)
- [11] 倪丽萍. 基于分形技术的金融数据分析方法研究[D]. 合肥: 合肥工业大学管理学院, 2010.
(Ni L P. Research on financial data analytical method based on fractal technology[D]. Hefei: School of Management, Hefei University of Technology, 2010.)
- [12] 陶熠昆. 基于非线性系统分析与动态复杂网络的医学数据分析与集成[D]. 杭州: 浙江大学电气工程学院, 2010.
(Tao Y K. Nonlinear systematic methods and dynamic complex networks for medical data analysis and integration[D]. Hangzhou: School of Electrical Engineering, Zhejiang University, 2010.)
- [13] 闫光辉, 李战怀, 党建武. 基于多重分形的聚类层次优化算法[J]. 软件学报, 2008, 19(6): 1283-1300.
(Yan G H, Li Z H, Dang J W. Finding natural cluster hierarchies based on multifractal[J]. J of Software, 2008, 19(6): 1283-1300.)
- [14] Belussi A, Faloutsos C. Estimating the selectivity of spatial queries using the correlation fractal dimension[C]. Proc of the 21th Int Conf on Very Large Data Bases(VLDB). Zurich, 1995: 299-310.
- [15] 刘海峰, 代正华, 陈峰, 等. 混沌动力系统小波变换模数的关联维数[J]. 物理学报, 2002, 51(6): 1186-1189.
(Liu H F, Dai Z H, Chen F, et al. Calculating the correlation dimension of dynamical system with wavelet analysis[J]. Acta Physica Sinica, 2002, 51(6): 1186-1189.)
- [16] 陆君安, 吕金虎, 陈士华. Chen's 混沌吸引子及其特征量[J]. 控制理论与应用, 2002, 19(2): 308-310.
(Lu J A, Lv J H, Chen S H. Chen's chaotic attractor and its characteristic quantity[J]. Control Theory & Applications, 2002, 19(2): 308-310.)
- [17] 刘景夏, 胡冰新. 一种基于小波包变换的关联维数计算方法[J]. 解放军理工大学学报: 自然科学版, 2006, 7(3): 229-231.
(Liu J X, Hu B X. Method of computing correlation dimension based on wavelet packet transform[J]. J of PLA University of Science and Technology: Natural Science Edition, 2006, 7(3): 229-231.)
- [18] 王忠林, 许明清. 一个新的超混沌系统分析与FPGA实现[J]. 重庆邮电大学学报: 自然科学版, 2010, 22(4): 445-449.
(Wang Z L, Xu M Q. Analysis and FPGA implementation of a novel hyperchaotic system[J]. J of Chongqing University of Posts and Telecommunications: Natural Science Edition, 2010, 22(4): 445-449.)

(责任编辑: 郑晓蕾)