

# 中国城市化水平的自回归与功率谱分析

陈彦光

(北京大学城市与环境学院 城市与区域规划系, 北京 100871)

**摘要:** 利用 1949~2000 年的城市人口比重数据, 开展中国城市化过程的自相关和功率谱分析, 建立了 ARMA(1,  $q$ ) 模型。中国的城市化过程具有 1 阶自相关特征: 上一年的一个变动直接影响下一年, 间接影响则可达 10 年之久。ARMA(1,  $q$ ) 模型表明中国的城市化过程在趋势性上附加有丰富的随机性。对提取趋势之后的序列进行功率谱分析, 发现趋势性和随机性的背后隐含一个长度为 30 年左右的周期波动。根据上述研究, 将中国的城市化过程分解为三种变动: 趋势性、周期性和随机性。研究结论对发展更为完善的城市化预测方法, 以及对指导具有中国特色的城市化建设, 都可能具有一定程度的启示意义。

**关键词:** 城市化; 城市变化; 城市人口比重; 自回归; 谱分析; 中国

**文章编号:** 1000-0585(2007)05-1021-12

## 1 前言

中国的区域城市化研究存在三大困难。一是缺乏可比的基础数据。城市形态是一种无标度现象<sup>[1]</sup>, 因此无法客观地界定城区范围, 从而无法客观地确定城市人口<sup>[2]</sup>。为此需要建立科学的城市地域定义标准。由于城市定义尚未规范, 加之统计方面的种种原因, 我国目前无法找到系统并且具有可比性的城市人口比重数据。二是缺乏有效的研究方法。城市化的动力学过程及其导致的空间结构都非常复杂, 很难采用单一的方法进行研究, 一种方法往往只能从一个角度反映问题。因此需要多种方法, 以便开展全方位的思考。三是对城市化过程及其动力学机制的基础性认识尚未深入。众所周知, 城市和区域都是复杂的空间系统<sup>[3,4]</sup>。这类系统具有反直观性, 而我们的判断通常限于感性认识。上个世纪后期, 一些新的研究方法如动态模拟技术引入城市化研究, 人们借助细胞自动机 (CA)、多重智能体 (MAS) 等方法模拟城市化的时空演变过程<sup>[5,6]</sup>, 为我们提供了新的认知视角。但是, 对于中国这种情况复杂的大国, 城市化研究需要更多的探索途径。

我国目前处于社会经济发展的重要转型时期<sup>[7,8]</sup>, 城市化是亚洲乃至世界关注的重大课题<sup>[9]</sup>。如何在受到上述三个方面局限的情况下, 对我国城市化进程进行演化分析和预测, 是一个不可忽视的学术任务。基础数据问题在短期内难以圆满解决, 我们只能充分利用现有的数据资料, 力求去伪存真。方法的问题虽非短期内可以完善, 我们有必要通过内涵和外延两种途径寻求对城市化动力学机制的深入认识。本文的目的, 就是要在研究方法上进行力所能及的延拓, 打开我国城市化探索的新的窗口。首先进行城市化时间序列的

收稿日期: 2006-10-08; 修订日期: 2007-05-23

基金项目: 国家自然科学基金重点资助项目(40335051); 美国 Urban China Research Network Small Grant Program 资助课题(2003Spring)的基础理论部分。

作者简介: 陈彦光 (1965-), 男, 博士, 副教授。主要从事城市和理论地理学研究, 重点研究地理分形和空间复杂性。

自相关分析, 并建立自回归—移动平均模型, 据此解释城市化过程的趋势性和随机性; 然后借助 Fourier 变换, 利用提取趋势后的时间序列开展功率谱分析。基于 Fourier 变换的功率谱分析在科学研究中具有不可替代的地位, 有人誉之为将复杂对象分解为基本要素的一种十拿九稳的方法, 该方法在人文地理研究中业已表现出自身的优势<sup>[10,11]</sup>。文章的研究目标之一是分解复杂的城市演化过程, 提供城市化进程分析的新的思路, 并且得出一些对我国城市化发展具有现实意义的研究结论。

## 2 趋势性与随机性: 城市化过程的自回归分析

### 2.1 数据资料和模型的初步估计

接下来利用全国性的城市人口比重数据开展进一步的分析(表1)。数据时段为1949~2000年, 其中1991~2000年的数据经过周一星、于海波的修补工作<sup>[12]</sup>, 修补的理由有关文献已经详细论述<sup>[12~14]</sup>。由于局部误差对宏观分析的研究结果影响并不明显, 如果采用国家统计局公布的数据, 分析结论没有本质的不同。但国家统计局公布的1990年后的数据问题比较明显, 而基于修补之后的数据可以得到更为准确的模型参数。

表1 中国城市人口比重的数据(1949~2000)(单位:%)

Tab. 1 The data of China's urbanization level 1949~2000 (unit: %)

年份	城市化水平	年份	城市化水平	年份	城市化水平	年份	城市化水平
1949	10.64	1962	17.33	1975	17.34	1988	25.81
1950	11.17	1963	16.84	1976	17.44	1989	26.21
1951	11.78	1964	18.37	1977	17.55	1990	26.62
1952	12.46	1965	17.98	1978	17.92	1991	26.98
1953	13.31	1966	17.86	1979	18.96	1992	28.68
1954	13.69	1967	17.74	1980	19.39	1993	29.63
1955	13.48	1968	17.62	1981	20.16	1994	30.57
1956	14.62	1969	17.50	1982	21.13	1995	31.44
1957	15.39	1970	17.38	1983	21.62	1996	32.23
1958	16.25	1971	17.26	1984	23.01	1997	33.28
1959	18.41	1972	17.13	1985	23.71	1998	34.24
1960	19.75	1973	17.20	1986	24.52	1999	35.23
1961	19.29	1974	17.16	1987	25.32	2000	36.26

说明: 周一星、于海波城市人口修补方案有二: 一是从三普(1982)到五普(2000); 二是从四普(1990)到五普(2000)。本文采用第二种方案的数据, 从1990年开始为修订后的结果, 1990年以前的数据与国家统计局公布的资料一样, 参见《中国人口统计年鉴》, 2001。

刻画城市、乡村人口时空转移与演化过程的动力学模型有多种形式, 包括 Keyfitz 模型、Rogers 模型、异速生长模型, 等等, 其中最著名也最有效的是联合国模型<sup>[15,16]</sup>。从联合国模型出发, 在一定条件下可以导出城市化过程的 Logistic 方程, 因此周一星将预测城市人口比重的 Logistic 模型也称为联合国模型<sup>[17]</sup>。无论根据时空动力学分析, 抑或根据理论上的考虑, Logistic 模型都是模拟城市化过程曲线的最佳方程, 但是, 并非任何一个国家或者区域都遵循这种演化模式。用中国的城市化水平数据拟合 Logistic 模型, 饱和值不收敛, 即模型

$$L_t = \frac{L_m}{1 + (100/L_0 - 1)e^{-bt}} \quad (1)$$

的饱和参数  $L_m > 100$ 。式中参数  $L_0$  表示初始年份的城市化水平， $b$  为初始增长率。这种结果是不正常的，这意味着中国的城市化进程存在某种不合理的加速问题。简而言之，如果 Logistic 模型拟合过程中出现饱和参数大于理论上的最大容量（100%）的现象，那就意味着研究对象不属于 Logistic 过程，不能采用 Logistic 模型进行描述。否则，模型参数违背事理，建模的结果没有意义。

进一步分析表明，中国城市化水平增长曲线有两个备选模型：

$$\text{一是时间的抛物线函数} \quad L_t = L_0 + at^2 \quad (2)$$

该模型可以较好地拟合城市化水平序列的五个控制点——五次人口普查数据。式中参数  $L_0$  表示初始年份的城市化水平， $a$  与增长率有关。

$$\text{其二是幂次双曲线模型} \quad \frac{1}{L_t} = \alpha + \beta t^\sigma \quad (3)$$

该模型可以较好地拟合 1949~2000 年的城市化水平数据，式中希腊字母  $\alpha$ 、 $\beta$ 、 $\sigma$  为参数。显然，式(2)和式(3)都不能用于中长期预测。

动力学分析表明，Logistic 方程隐含着—个环境约束参量，是人地协调型增长；抛物线函数是一种无约束型增长，缺乏限制增长过程的环境参量；幂次双曲线则是一种资源掠夺型增长——增长参量包含有破坏环境的信息。后面两种情况都属于非节约型增长，都会因为过度地加速而导致系统的崩溃。由此可见，式(2)和式(3)都是不可持续增长模式。人类社会作为一种自组织系统，城市化过程作为一种自组织过程<sup>[3,18]</sup>，当然不会允许有害模式为时太久。事实上，由于科学发展观的提出，我国目前正在开拓新型的工业化道路。可以想见，新型的工业化将会导致新型的城市化道路。

不论怎样，确定性模型的存在，都暗示一种趋势性的存在。中国的城市化存在明确的增长的趋势，这也符合城市化过程的一般规律。

借助随机序列的时域分析和频域分析，可以揭示中国城市化过程在趋势以外的变动特征。时域分析包括自相关分析和自回归—移动平均分析，频域分析则主要是基于 Fourier 变换的谱分析——谱分析用于时间序列就是功率谱分析<sup>[10]</sup>，用于空间序列就是波谱分析<sup>[11]</sup>。有关方法在地理学领域早有应用<sup>[19]</sup>，读者也可以从有关地理学词典如 R. J. Johnston 的《人文地理学词典》<sup>[20]</sup>中查阅一些相关的概念。

## 2.2 AR(2)过程与趋势性

对于时间序列的时域分析和预测，有三种基本模型：其一是自回归(Auto-Regression, AR)，其二是移动平均(Moving Average, MA)，其三是自回归—移动平均(Auto-Regressive Moving Average, ARMA)。自回归、移动平均等概念最初由苏联统计学家 E. Slutsky 和英国统计学家 G. U. Yule 提出。G. E. P. Box 等在此基础上发展了一套完善的建模和预测规范<sup>[21, 22]</sup>。因此，移动平均—自回归模型通常也称为 Box-Jenkins 模型。关于 MA、AR 和 ARMA 建模，读者可以参考 Box 等的英文影印本<sup>[21]</sup>或者中译本<sup>[22]</sup>了解相关概念和具体方法。下面给出基于自相关分析的自回归模型，即 AR 过程。

序列的自相关性关系到模型预测的可靠性。如果一个时间序列存在很强的自相关性，则以时间为自变量建立的常规预测模型如式(1)~式(3)不再有效——这类模型可以作为趋势判断的参考，但不能作为发展预测的工具。原因在于此时模型的残差序列不再是白噪声序列，或者说不再服从正态分布假设。因此，模型参数的精度较低，从而趋势外推的置信范围过宽。在这种情况下，对时间序列本身开展自相关分析，并建立自回归模型即

AR( $p$ )过程, 可以为我们提供系统演化的有用信息。

中国城市化水平的自相关函数(ACF)和偏自相关函数(PACF)的特征如图1所示。

第一, ACF特征介于单边衰减和阻尼震荡衰减之间, 暗示中国城市化的自回归(AR)的阶次 $p$ 可能介于1~2之间。在 $p=11$ 的时候, ACF进于2倍的标准误差带内, 表明自相关过程的间接影响可达10年左右。剩余的微弱影响甚至达到15~16年之久, 这意味着有可能隐含一个30~32年的周期波动(图1a)。

第二, PACF表现出明确的1阶截尾, 暗示中国的城市化可能为1阶自回归即AR(1)过程(图1b)。

首先利用城市化水平的时序序列(表1), 借助普通最小二乘法(OLS), 估计出1阶自回归模型

$$L_t = -0.086 + 1.029L_{t-1} + \varepsilon_t \quad (4)$$

拟合优度为 $R^2 = 0.99$ 。回归系数大于

1但近似为1, 根据经典的DF检验可知, 在这种情况下, 应该采用AR(2)过程<sup>[23]</sup>。2阶自回归的统计参量见表2。AR(2)过程如下

$$L_t = -0.012 + 1.406L_{t-1} - 0.390L_{t-2} + \varepsilon_t \quad (5)$$

实际上估计AR模型参数的方法有多种, OLS方法是一个经典的框架。作为参照, 借助

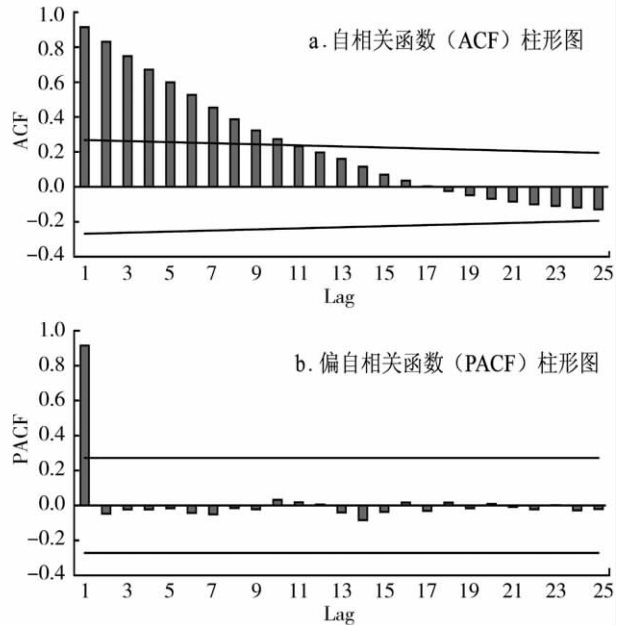


图1 中国城市化水平的相关函数柱形图

Fig. 1 Histograms of autocorrelation function (ACF) and partial autocorrelation function (PACF) of China's urbanization level time series (a. ACF histogram; b. PACF histogram)

表2 中国城市化水平2阶自回归过程的最小二乘估计

Tab. 2 The 2nd order autoregression results of China's level urbanization based on ordinary least squares (OLS) estimator

回归统计		方差分析	df	SS	MS	F	Significance F
Multiple R	0.996	回归分析	2	2011.626	1005.813	2605.378	0.000
R Square	0.991	残差	47	18.144	0.386		
Adjusted R Square	0.991	总计	49	2029.770			
R Square 标准误差	0.621						
Durbin-Watson	1.843						
观测值	50						
	Coefficients	标准误差	t Stat	P-value	下限 95.0%	上限 95.0%	
Intercept	-0.012	0.311	-0.038	0.970	-0.638	0.615	
$x_{t-1}$	1.406	0.135	10.432	0.000	-0.670	-0.110	
$x_{t-2}$	-0.390	0.139	-2.801	0.007	1.135	1.677	

说明: 利用电子表格 Excel、统计分析软件 SPSS 或者数学软件 Mathcad 等均得到表中的结果。

Prais-Winsten 估计法得到如下模型

$$L_t = 0.002 + 1.254L_{t-1} - 0.236L_{t-2} + \epsilon_t \quad (6)$$

该模型对应的 DW 值=1.914, 这表明自回归模型消除了序列相关的影响, 但 2 阶自回归系数(ARC)的  $P$  值达到 0.119, 置信度低于 90%, 故采用 OLS 方法的估计结果。

进一步地, 估计 3 阶、4 阶模型即 AR(3)和 AR(4)过程。结果表明, 2 阶以上的模型参数或多或少存在问题, 由此可以初步判断, 2 阶 AR 模型最为可靠。一个间接的判据是: 如果一个时间序列是一种 AR(2)过程, 则可利用它的 1 次差分序列建立 AR(1)过程。原因是: 对一个 1 阶 AR 过程求微分, 然后将变量离散化, 总可以得到一个 2 阶 AR 过程。虽然由于离散—连续变换常常导致一些参数误差, 但不影响对整个系统性质的判断。

中国城市化水平的 1 次差分序列(即单整  $I=1$  的序列)如图 2 所示, 这个序列的趋势性不再明显。PACF 为 1 阶截尾, 与 AR(1)过程一致(图 3)。基于差分序列建立 AR(1)过程

$$\Delta L_t = 0.286 + 0.439\Delta L_{t-1} + \epsilon_t \quad (7)$$

转换为原始序列表示的 AR(2)过程便是

$$L_t = 0.286 + 1.439L_{t-1} - 0.439L_{t-2} + \epsilon_t \quad (8)$$

这个模型与前面直接估计的模型即式(5)非常接近, 这表明, AR(2)过程符合中国城市化过程的实际, 可以用它模拟中国城市人口比重的变化曲线。

2 阶 AR 模型与原始序列的 PACF 似乎存在矛盾, 但与其 ACF 的特征存在对应关系。为了解决上述矛盾问题, 有必要引入移动平均即 MA 过程开展分析。

### 2.3 ARMA(1,q) 过程与随机性

对一次差分序列开展自相关分析, 发现 ACF 和 PACF 均为 1 阶截尾, 但 ACF 震荡衰减, 具有 MA( $q$ )特征(图3)。基于差分序列的 AR(1)过程的回归系数小于 1, 表明模型可逆且平稳, 转换为移动平均过程可得 MA( $q$ )模型如下

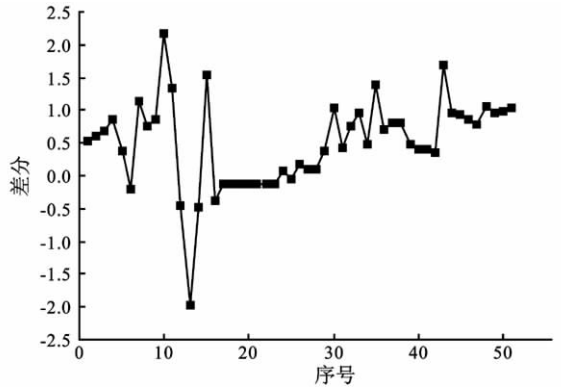


图 2 中国城市化水平的差分序列变化特征(1949~2000)

Fig. 2 Difference of time series of China's urbanization level (1949~2000)

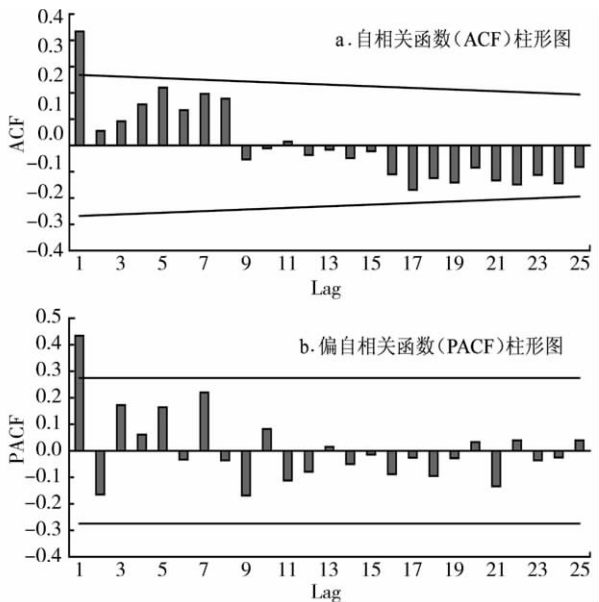


图 3 中国城市化水平差分序列的相关函数柱形图

Fig. 3 Histograms of autocorrelation function (ACF) and partial autocorrelation function (PACF) of China's urbanization level difference series (a. ACF histogram; b. PACF histogram)

$$\begin{aligned}
 \Delta L_t &= 0.286 \left( \sum_{i=0}^p 0.439^i \right) + 0.439^p \Delta L_{t-p} + \sum_{j=0}^q 0.439^j \varepsilon_{t-j} \\
 &= 0.286 \left( \frac{1}{1-0.439} \right) + \sum_{j=1}^q 0.439^j \varepsilon_{t-j} \\
 &= 0.510 + \sum_{j=1}^q 0.439^j \varepsilon_{t-j}
 \end{aligned} \tag{9}$$

模型常数 0.510 非常接近差分序列的平均值 0.503, 这与理论的推断<sup>[24]</sup>一致。将模型一般化就是

$$\Delta L_t = \mu + \sum_{j=0}^q \varphi^j \varepsilon_{t-j} \tag{10}$$

式中  $q$  理论上趋于无穷大。将差分还原得到 ARMA(1,  $q$ ) 模型

$$L_t = \mu + L_{t-1} + \lim_{q \rightarrow \infty} \sum_{j=1}^q \varphi^j \varepsilon_{t-j} \tag{11}$$

可见中国的城市化过程是由 1 阶 AR 和高阶乃至无穷阶 MA 过程复合而成——MA 代表一个随机扰动过程。

至此可以给出自回归分析的基本结论: (1) 在理论上, 中国的城市化过程是一个 1 阶自回归加上无穷阶移动平均过程。1 阶自回归暗示城市化过程的一年滞后影响, 无穷阶移动平均表示城市化过程存在较多的不确定性因素。二者的叠加效应是 2 阶滞后的假象。(2) 在现实中, 中国城市化过程的直接自相关主要是上一年影响下一年, 这个“局部”性质 (locality) 与随机性叠加为一个特征丰富的变化过程。如果考虑时间上的关联效应, 则一个年份的某种影响可以传递到 10 年之后。(3) 自相关过程暗示, 采用常规的回归模型预测城市化水平的增长, 其结果准确性较低, 有必要结合自回归模型进行综合判断。

### 3 周期性: 城市化过程的功率谱分析

#### 3.1 周期性的谱识别

上面的建模结果表明, 中国城市化的动态过程丰富而又复杂, 有必要分析趋势性背后的演化行为。我们知道, 复杂的时间序列通常包含有随机性、趋势性、季节性和周期性等多种行为模式, 必要时可以进行季节分解或者滤波。对于城市化时间序列, 不太可能存在季节性, 但有可能存在某种周期行为。根据 Berry 的研究可知, 经济上的长波节律对全球城市增长存在明显的影响, 经济发展的长波历史格局甚至与城市增长的空间格局具有动态联系<sup>[25]</sup>。不仅如此, 在 Kondratiev 的长波变动过程中还存在较短时期的周期振荡乃至更小时段的随机涨落<sup>[26]</sup>。经济上不同尺度的波动行为都可能在城市化过程中留下烙印, 问题在于如何揭示政治经济发展与城市化的互动关系的后果。

对于存在明确趋势并且可以据之建模的序列, 可以不必进行滤波, 而是借助趋势模型分解序列。对于本例, 提取时间序列趋势的方式有三种: 一是提取线性趋势, 这是常规方法; 二是提取 Logistic 趋势, 这是理想方法, 三是提出最佳拟合趋势 (抛物线或者幂次双曲线), 这是现实方法。本文采用第二种方法, 亦即提取 Logistic 趋势。这么做的原因有二: 其一, 无论提取哪种趋势, 最后的分析结果没有本质的区别; 其二, 从理论看和长期看, 中国的城市化过程必将回归到 Logistic 模式。借助表 1 中的数据估计 Logistic 模型

$$\hat{L}_t = \frac{100}{1 + 7.387e^{-0.0243t}} \tag{12}$$

拟合优度为  $R^2 = 0.882$ 。据此估算残差

$$e_t = L_t - \hat{L}_t \tag{13}$$

残差曲线如图 4 所示，这种变动可能具有某种周期规律。

接下来利用快速 Fourier 变换 (Fast Fourier Transform, FFT) 对残差进行谱分析。Fourier 分析是积分变换的一种常规方法，最早由法国数学家、物理学家 Joseph Fourier (1768~1830) 提出<sup>[27]</sup>，在科学研究中有着广泛的应用<sup>[28]</sup>。Fourier 分析现已成为频谱分析的基本工具，这种工具早就被引入地理分析<sup>[19,20]</sup>。中国城市化水平残差的谱分析结果表明，的确存在一个周期变化特征，在频率  $f=0.03125$  的地方出现一个明显突出的峰值，其倒数即为周期长度  $T=1/f=32$ 。不难根据调和分析的 Fisher 统计量检验表检验周期的是否成立。在 FFT 过程中，序列长度  $s=64/2=32$ ，检验周期数目  $r=1$ 。在显著性水平  $\alpha=0.05$  下查得临界值为

$$g_c = Z_{\alpha}(s, r) = Z_{0.05}(32, 1) = 0.188$$

根据频谱分析结果，计算 Fisher 统计量得到

$$g = \frac{\max_{0 \leq f \leq 0.5} (P(f))}{\sum_{f=0}^{0.5} P(f)} = \frac{69.029}{121.616} = 0.568 > g_c = 0.188$$

检验通过，这意味着有 95% 的把握相信周期性成立。至此得出如下结论：(1) 在长期增长的趋势背后，中国的城市化水平隐含一个周期波动，波长约为 32 年。考虑到 FFT 算法引起的误差，估计周期长度为 30 年左右。(2) 从理论上讲，如果城市化水平按照 Logistic 过程增长，则从模型即式(12)的参数值判断，不可能出现周期波动。因此，要么中国的城市化过程不是 Logistic 增长，要么还有导致周期的外在因素，甚至兼而有之。(3) 中国城市化过程的这种周期变动可能对单个城市人口增长具有很大的影响，或者各个城市的发展过程与中国的城市化过程内在相关，因为具体城市如北京、天津的成长过程背后也隐含有这种周期变动<sup>[10]</sup>。

### 3.2 周期性的谱解释

对上述周期现象一个可能的解释是国家政治经济层面的变化。从 1949 年解放至今已经半个多世纪，其中 1978~1981 年是中国社会转型期的开端，自 1949 年到转型，时段大约 30 年；此前的“文革”爆发和知识青年的上山下乡运动，此后的计划经济向社会主义市场经济转轨，都可能成为触发城市化进程波动的某种契机。虽然上述因素可能引起起伏，但是城市化过程自身的动力学机制理当更有影响。或者，内因和外因交互作用。下面重点分析内因问题。利用原始时间序列进行谱分析，建立频谱关系

$$P(f) \propto f^{-\beta} \tag{14}$$

估计模型为

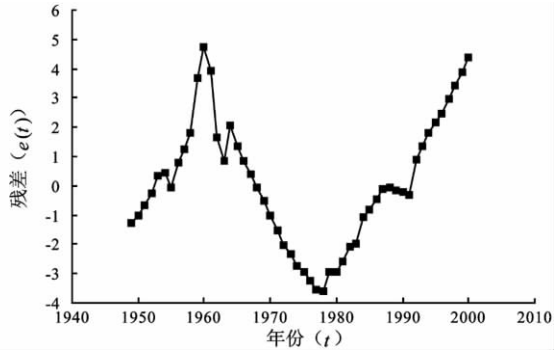


图 4 基于 Logistic 曲线的中国城市化水平的残差变化图 (1949~2000 年)

Fig. 4 Residual curve of China's urbanization level based on the logistic model (1949~2000)

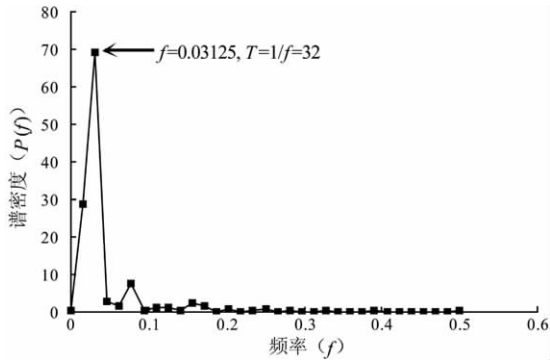


图5 中国城市化水平的 logistic 残差频谱图

Fig. 5 Relationship between frequency and spectrum density of residual series based on the logistic model of China's urbanization level

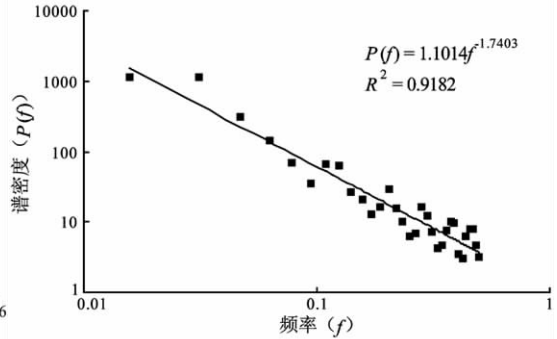


图6 中国城市化水平时间序列的频谱图(1949~2000)

Fig. 6 Relationship between frequency and spectrum density of China's urbanization level time series

$$P(f) = 1.101f^{-1.74} \quad (15)$$

拟合优度为  $R^2=0.918$  (图6)。谱指数为  $\beta=1.74$ , 这表明城市化过程是一种广义的  $1/f$  噪声<sup>[29]</sup> (如果采用国家统计局数据, 参数稍有区别, 但规律不变<sup>[18]</sup>)。根据关系<sup>[30]</sup>

$$\beta = 2H + 1 = 5 - 2D \quad (16)$$

算得 Hurst 指数  $H$  和分维数  $D$  为  $H=0.37, D=1.63$ 。于是 2 阶自相关系数为负值  $R(2) = 2^{2H-1} - 1 = -0.165$ 。这表明, 过去的一个增加量意味着未来的一个减少量, 因此必然引起波动。基本结论如下: 根据功率谱指数以及据此转换得出的 Hurst 指数, 中国的城市化过程具有反持久性——上一个时段的增量导致下一个时段的减量, 反之亦然。如果按照目前的模式和速度, 预计若干年后还会出现城市化水平的背景性起伏变化。

更深入的探讨表明, 中国城市化水平的周期波动还可能与中国城乡人口迁移和演化的时空动力学特征有关(限于篇幅, 有关问题将另文专论)。

## 4 复杂性: 城市化的多重运动叠加模型

### 4.1 中国城市化的三种变动

根据上述定量和定性分析的结果, 中国城市化过程可以分解为如下三种变动:

“趋势性+周期性+随机性”

趋势性为首要特性, 表现为双曲线或者抛物线, 从长远看应该接近 Logistic 曲线。城市化为何具有趋势性, 早就有比较公认的解释<sup>[31]</sup>, 至于为什么一个国家或者地区形成一种特有的趋势, 则是需要开展动力学分析的问题, 但这不是本文的主要任务。周期性或波动性是在时间序列中剔除趋势性之后, 通过谱分析体现出来。周期性是附加于趋势性的, 将趋势性剔除之后, 可以发现这种周期性。周期长度大约为 30 年。随机性通过残差或者 ARMA 过程表示出来。随机性伴随一个区域的城市化全程, 无论剔除趋势与否, 都可以发现这个随机性的存在。趋势性反映了城市化过程的主导力量, 周期性表明积极和消极两种因素的对抗。趋势性和周期性共同暗示了中国城市化过程的动力学特征。至于随机性, 则是最为丰富多彩的一个方面, 系统的复杂性往往借助随机性变相地表现出来。随机性反映了城市化过程的不可控制和不可预测因素, 或者叫做导致突发变动的机遇因子。由于机遇因子以及相关的



机制,城市化演化过程实际上是一种自组织过程<sup>[18,32]</sup>,城市管理理当考虑自组织思想。

#### 4.2 中国城市化预测的难题

上述分析除了揭示中国城市化过程的一些基本规律之外,还引发如下难题:如何预测未来中国城市化水平的增长趋势?我们知道,最基本的数学分析方法是,利用因果关系寻求解释,或者借助某种现象随时间的变化进行趋势外推,从而得到所谓预测结果。对于第二种情况,时间变量与因变量之间没有因果关系,故有人称之为时间虚拟变量(time dummy)<sup>[24]</sup>——这是不同于分类变量的一种哑变量。无论属于那种情况,回归分析都有一个基本的要求,那就是序列之间没有明确的自相关性,否则分析结论和预测结果都不可靠。上述自相关和自回归分析表明,中国的城市化过程具有自相关性,这种自相关性会在常规回归模型的残差序列中表现出来。在这种情况下,常规的回归分析不可能给出有效的预测结果,必须在模型中引入 AR 过程。

通常的 AR 分析是基于线性趋势假设的,而城市化水平增长显然是非线性的,这就要求我们建立更为复杂的非线性自回归分析模型。方法是:(1)借助非线性回归方法寻找合适的非线性常规模型;(2)利用这个模型将时间序列线性化;(3)基于线性化时间序列建立自回归模型;(4)借助常规非线性模型将普通自回归模型转换成非常规的非线性自回归模型;(5)利用非线性自回归模型进行预测分析。

然而,上述方法对于中国的城市化水平预测却不是有效的途径,困难在于无法建立中国城市化水平增长的非线性数学模型。Logistic 模型是理论意义的,但中国城市化水平的时序在整体上并不服从这个模型。双曲线和抛物线模型

可以更好地拟合城市化水平序列,但这两个序列都是“爆炸”型的,时间序列会因为不断加速增长而趋于崩溃,因此根本不可以用于长期预测。由于找不到恰当的常规非线性模型作为数据线性化的转换工具,也就无法建立非线性自回归预测模型。作为参考,表 3 给出了基于常规 Logistic 曲线和线性自回归分析的预测数据。这些线性自回归分析可以用于近期估计,但对于远期预测可能出现荒谬的结果。

城市化增长过程理当服从 Logistic 模型。从动力学特征看来,Logistic 增长模型有一个环境约束的减速因子。但是,借助中国的城市化水平建立 Logistic 模型,饱和值却不收

表 3 中国城市化水平的 Logistic 预测和线性自回归估计:2001~2010 年(单位:%)

Tab. 3 Prediction of China's urbanization level using logistic model and auto-regression method (unit: %)

年份	常规非线性		线性自回归			统计值
	Logistic 预测	预测 1	预测 2	预测 3		
2001	36.88	37.21	37.16	37.00	37.66	
2002	37.80	38.15	38.04	37.61	39.09	
2003	38.71	39.10	38.94	38.16	40.53	
2004	39.63	40.06	39.85	38.69	41.76	
2005	40.55	41.05	40.78	39.21	未知	
2006	41.47	42.06	41.74	39.72	未知	
2007	42.39	43.10	42.72	40.23	未知	
2008	43.30	44.16	43.72	40.74	未知	
2009	44.21	45.24	44.75	41.26	未知	
2010	45.12	46.36	45.80	41.77	未知	

说明:①Logistic 预测参见文献<sup>[33]</sup>的说明。②线性自回归分别采用三个模型:预测 1 采用基于 OLS 法的 AR(2)模型,即式(5);预测 2 采用基于 Prais-Winsten 法的 AR(2)模型,即式(6);预测 3 采用基于差分序列和 OLS 法的 AR(2)模型,即式(8)。③统计数据来自《中国统计年鉴》。根据国家统计局的资料,2004 年底,中国城市化水平是 43%(参见 <http://www.hexun.com>,年鉴上有所调整。现在又有人估计 2005 年的城市化水平为 43%,看来中国城市化水平的实际值要到下一次人口普查之后才能知道究竟。

敛。而且,如果城市化水平增长服从 Logistic 模型,则当城市化水平达到饱和值的一半时一定减速<sup>[33]</sup>。但现实的情况并非如此。实际上,如前所述,中国的城市化过程并不服从标准的 Logistic 模型,只有 1982 年后经过文献[12]修补过的数据具有 Logistic 趋势,原因在于赖以修补的联合国模型本身就是 Logistic 方程。从动力学上分析,如果一个国家或者区域的城市化过程服从抛物线或者双曲线模型,就表明城市化速度可能失控,城市人口增长的过程可能伴随有比较严重的资源掠夺过程。中国的情况究竟如何,需要进行大量的实地考察才能明确。令人欣慰的是,由于近年来中央政府采取了强有力的措施推行科学发展观,许多地方都在实行节能减排、集约生产,不正常的增长模式将会得到纠正,城市化也可望走上健康、合理的轨道。

## 5 结束语

中国的城市化在改革开放之后逐步进入了新的阶段<sup>[34]</sup>,对中国城市化过程有多种理论解释<sup>[35]</sup>,但很多观点见仁见智,言人人殊。揭示中国城市化过程的动力学特征及其演化规律是一件非常困难的事情,需要很多基础工作的支持,包括数据、理论和技术基础等诸多方面。本文着重在理论和方法方面开展一些基础性工作的尝试,这些工作虽然初步,但对进一步认识中国城市化的特征可能会有一些有益的启示。文章的要点概括如下。

第一,通过自相关、自回归和功率谱分析,可将中国的城市化过程分解为三种基本变动:一是趋势变动,备选模型有 Logistic 模型、抛物线模型和幂次双曲线模型;二是周期波动,周期长度大约 30 年;三是随机变动,受到多种因素的影响,内容最为丰富,且与城市变化的自组织过程有关。解析上述三种变动需要开展更为深入的探讨工作,限于篇幅,留待今后专文论述。

第二,通过自相关和自回归分析发现,中国的城市化过程具有 1 阶自相关特征:上一年的一个变动直接影响下一年——这暗示一种演化过程的局域性质,但间接影响可达 10 年之久。因此城市化的政策务必谨慎,一个失误可能在 10 年之后出现不良反应,借助周期波动,其负面效应可能传递得更远。

第三,通过谱分析和基本趋势分析发现,中国近 50 年的城市化过程有两大数理“特色”:一是不稳定性,即反持久性,表现为周期波动;二是不可持续性,即非节约型增长,表现为幂次双曲线增长模式以及 Logistic 饱和参数的不收敛。这是宏观特色之一。如何将中国的城市化水平引导到 Logistic 的轨道上来,是中国政府未来任重而道远的工作。

致谢:作者衷心感谢两位评审专家对本文修改的建设性意见。一位专家对文章的关键性措辞进行了把关,使得本文的语言表达得以改善。

## 参考文献:

- [1] Batty M, Longley P A. Fractal Cities: A Geometry of Form and Function. London: Academic Press, 1994.
- [2] 陈彦光. 中国的城市化水平有多高? 城市规划, 2003, 27(7): 12~17.
- [3] Allen. P M. Cities and Regions as Self-Organizing Systems; Models of Complexity. Amsterdam: Gordon and Breach Science Pub., 1997.
- [4] Wilson A G. Complex Spatial Systems; The Modelling Foundations of Urban and Regional Analysis. Singapore: Pearson Education Asia Pte Ltd., 2000.

- [ 5 ] Clarke K C, Hoppen S, Gaydos L. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design*, 1997, 24: 247~261.
- [ 6 ] Sanders L, Pumain D, Mathian H, *et al.* SIMPOP: A multiagent system for the study of urbanism. *Environment and Planning B: Planning and Design*, 1997, 24(2): 287~305.
- [ 7 ] 冯健. 转型期中国城市内部空间重构. 北京: 科学出版社, 2004.
- [ 8 ] Friedmann J. *China's Urban Transition*. Minneapolis, MN: University of Minnesota Press, 2005.
- [ 9 ] Ness G D, Talwar P P (eds.). *Asian Urbanization in the New Millennium*. Singapore: Marshall Cavendish Academic, 2005.
- [ 10 ] 陈彦光, 刘继生. 基于引力模型的城市空间互相关和功率谱分析. *地理研究*, 2002, 21(6): 742~752.
- [ 11 ] 陈彦光, 刘继生. 城市人口分布空间自相关的功率谱分析. *地球科学进展*, 2006, 21(1): 1~9.
- [ 12 ] 周一星, 于海波. 以“五普”数据为基础对我国城镇化水平修补的建议. *统计研究*, 2002, (4): 44~47.
- [ 13 ] Zhou Y X, Ma L J C. China's urbanization levels: Reconstructing comparable time-series data based on the fifth population census. *China Quarterly*, 2003, 173: 184~204.
- [ 14 ] Zhou Y X, Ma L J. China's urban population statistics: A critical evaluation. *Eurasian Geography and Economics*, 2005, 46(4): 272~289.
- [ 15 ] United Nations. *Patterns of Urban and Rural Population Growth*. New York: U. N. Department of International Economic and Social Affairs, Population Division, 1980.
- [ 16 ] Karmeshu. Demographic models of urbanization. *Environment and Planning B: Planning and Design*, 1988, 15(1): 47~54.
- [ 17 ] 周一星. *城市地理学*. 北京: 商务印书馆, 1995.
- [ 18 ] 陈彦光. 城市化: 相变与自组织临界性. *地理研究*, 2004, 23(3): 301~311.
- [ 19 ] Bennett R J. *Spatial Time Series*. London: Pion, 1979.
- [ 20 ] Johnston R J 主编. 柴彦威, 等译. *人文地理学词典*. 北京: 商务印书馆, 2004.
- [ 21 ] Box G E P, Jenkins G M, Reinsel G C. *Time Series Analysis Forecasting and Control (3<sup>rd</sup> ed.)*. Beijing: Posts & Telecom Press, 2005.
- [ 22 ] Box G E P, Jenkins G M, Reinsel G C 著. 顾岚, 等译. *时间序列分析: 预测与控制*. 北京: 中国统计出版社, 1997.
- [ 23 ] Watsham T J, Parramore K. *Quantitative Methods in Finance*. London/Boston: International Thomson Business Press, 1997.
- [ 24 ] Diebold F X. *Elements of Forecasting (3<sup>rd</sup> ed.)*. Mason, Ohio: Thomson/South-Western, 2004.
- [ 25 ] Berry B J L. *Long-Wave Rhythms in Economic Development and Political Behavior*. Baltimore: Johns Hopkins University Press, 1991.
- [ 26 ] Berry B J L, Kim H. Long waves 1790-1990: Intermittency, chaos, and control. In: Kiel L D, Elliott E (eds.). *Chaos Theory in the Social Science: Foundations and Applications*. Ann Arbor: The University of Michigan Press, 1996. 215~236.
- [ 27 ] Fourier J 著. 桂质亮译. *热的解析理论*. 武汉: 武汉出版社, 1993.
- [ 28 ] Bracewell R N. *The Fourier Transform and Its Applications (3<sup>rd</sup> ed.)*. New York: McGraw-Hill Companies, Inc., 2000.
- [ 29 ] Bak P. *How Nature Works: The Science of Self-organized Criticality*. New York: Springer-Verlag, 1996.
- [ 30 ] Feder J. *Fractals*. New York: Plenum Press, 1988.
- [ 31 ] Levy J M. *Contemporary Urban Planning (7<sup>th</sup> ed.)*. Upper Saddle River, NJ: Pearson/Prentice Hall, 2006.
- [ 32 ] 陈彦光. 中国城市发展的自组织特征与判据. *城市规划*, 2006, 30(8): 24~30.
- [ 33 ] 陈彦光, 罗静. 城市化水平与城市化速度的关系探讨——中国城市化速度和城市化水平饱和值的初步推断. *地理研究*, 2006, 24(6): 1063~1072.
- [ 34 ] 周一星, 曹广忠. 改革开放 20 年来的中国城市化进程. *城市规划*, 1999, 23(12): 8~13.
- [ 35 ] 薛凤旋, 蔡建明. 研究中国城市化理论学派述评. *地理研究*, 1998, 17(2): 208~216.

# Modeling the urbanization process of China using the methods based on auto-correlation and spectral analysis

CHEN Yan-guang

(Department of Geography, Peking University, Beijing 100871, China)

**Abstract:** China's urbanization cannot be modeled by the logistic equation, which is followed by the USA's urbanization process. In order to reveal the features and property of China's urbanization, the auto-correlation and spectral analysis are employed to make a multifold study on time series of urbanization from 1949 to 2000. (1) An autocorrelation analysis is implemented, and partial autocorrelation function (PACF) has a first order cut-off. This implies that the urbanization process of China possesses a locality: a change in the  $i$ -th year only affects that in the  $(i+1)$ th year directly, but cannot affect the changes in and after the  $(i+2)$ th year. However, the auto-correlation function (ACF) suggests that a change perhaps influence a change ten years later indirectly. (2) An autoregressive analysis is made and an autoregressive moving-average (ARMA) model is built such as

$$L_t = \mu + L_{t-1} + \lim_{q \rightarrow \infty} \sum_{j=0}^q \varphi^j \varepsilon_{t-j} = 0.510 + L_{t-1} + \lim_{q \rightarrow \infty} \sum_{j=0}^q 0.439^j \varepsilon_{t-j}$$

where  $L_t$  is the  $i$ -th year's urbanization level,  $\varepsilon$  is an innovation or "random shock" (white noise),  $\varphi$  is a parameter, and  $q$  the order of moving average. (3) A spectral analysis is made based on the residuals of the logistic model, that is, the logistic trend of urbanization level is removed from the time series, and the result shows that there exists a periodic change behind the trend change. The wavelength (cycle length) is about 30 years. The Hurst exponent of the urbanization data is estimated to interpret the periodic behavior. The value of the Hurst exponent,  $H=0.37$ , suggests anti-persistence in the urbanization process of China. Based on the above analyses, the process of urbanization is divided into three parts: random process, periodic process, and trend process. Among the three different components of change in urbanization, trend is a basic process, cycle is an accessorial process, and random change is a complex process. The future of China's urbanization is hard to be predicted using the common methods because of auto-correlation and random disturbance, so new approaches should be found to conduct a convincing prediction.

**Key words:** urbanization; urban changes; urbanization level; auto-regression; power spectral analysis; China