

## 三江平原地下水埋深时间序列的混沌研究<sup>\*</sup>

付 强,李国良

(东北农业大学 水利与建筑学院,哈尔滨 150030)

**摘 要:**三江平原是我国最大的淡水沼泽区和重要的商品粮生产基地,地下水资源是该地区发展井灌水稻极为宝贵的源泉,但同时地下水系统又是一个复杂的非线性系统。首先,应用功率谱分析法对该地区地下水埋深序列进行定性分析;然后,在相空间重构的基础上,通过计算特征量 Kolmogorov 熵和关联维,得出三江平原地下水埋深序列中存在明显的混沌特征;最后,应用基于关联度的局域加权线性回归预测法进行混沌预测研究,得出比较理想的预测结果,为应用混沌理论进行三江平原地下水埋深变化规律研究提供依据。

**关键词:**混沌研究;功率谱分析;相空间重构;特征量;局域加权线性回归预测

**中图分类号:**P332.3

**文献标识码:**A

**文章编号:**1005-3409(2008)03-0031-04

## Chaos Studies on Series of Groundwater Depth in Sanjiang Plain

FU Qiang, LI Guo-liang

(College of Water Conservancy and Civil Engineering, Northeast Agricultural University, Harbin 150030, China)

**Abstract:** It is the largest swamp area of freshwater and the important production base of commodity food grain in China. The groundwater resource plays an important role in the development of well-irrigation paddy. The groundwater system is one kind of complicated nonlinear system. Firstly, the series of groundwater depth in sanjiang plain were analyzed qualitatively through power spectral analysis. Secondly, the conclusion that there is obvious chaos characteristic in the series of groundwater depth in Sanjiang plain was drawn through calculating the characteristic quantities, which are the Kolmogorov entropy and the correlative dimension of groundwater depth series, based on the reconstruction of state space. At last, the method of local adding-weight linear regression forecasting based on degree of incidence was applied to the series and the ideal results were got. The theoretical evidence would be provided for applying the chaos theory to study the variable law of groundwater depth in Sanjiang plain.

**Key words:** chaos studies; power spectral analysis; reconstruction of state space; characteristic quantity; local adding-weight linear regression forecasting

在地下水埋深的分析预测中,主要是以线性理论为基础,应用确定性数学模型和随机统计方法进行研究,预测精度不高<sup>[1]</sup>。近年来以混沌理论为代表的非线性预测方法的发展为研究地下水埋深提供了新的思路<sup>[2]</sup>。

地下水资源是三江平原广大井灌水稻区赖以发展水稻种植的极为宝贵的源泉。但是长期不合理、无节制的过量开采水资源,致使地下水资源遭到严重的破坏<sup>[1]</sup>。应用混沌理论,仅从地下水埋深的时间序列出发,通过相空间重构技术分析三江平原地下水埋深的变化规律性质,可以为进一步准确预测该地区地下水埋深提供理论依据。

### 1 研究区概况

三江平原地区是指黑龙江、松花江、乌苏里江汇流的三角地带和穆棱河流域,即西起小兴安岭,东至乌苏里江,北邻黑龙江,南达兴凯湖地区,是我国最大的淡水沼泽湿地地区。

这些年通过打井种稻,现在已经成为我国重要的商品粮生产基地。建三江分局始建于 1969 年,位于三江平原的东北部,地跨富锦、同江、饶河、抚远 4 个市县,西以青龙河与同江市为界,西南与富锦市相邻,东与东北与抚远县接壤,南以挠力河与饶河县为邻,东部及北部与俄罗斯隔江相望。

建三江分局地下水开发程度较高,地下水用量占总用水量的 99%,农灌用地下水占地下水总开采量的 98%,地下水资源已成为制约该局今后发展的重要因素。分局所属的创业农场超采严重,洪河农场与洪河自然保护区(国家级保护区,国际重要湿地)接壤,以创业农场和洪河农场的长观井资料(表 1)进行研究具有重要的代表性和理论价值。

### 2 功率谱分析

可以直接测量的数据包括时间序列的功率谱。功率谱密度函数(简称功率谱)是指单位频率上的能量,能反映其强

<sup>\*</sup> 收稿日期:2007-07-20

基金项目:国家自然科学基金项目(30400275);黑龙江省青年科学基金项目(QC04C28)

作者简介:付强(1973-),男,教授,博士生导师,主要从事水资源系统分析及农业系统工程建模与优化技术研究。E-mail:fuqiang@neau.edu.cn

度在频率上的分布情况,直观地揭示离散数据系列的周期性,是分析时间序列的常用方法<sup>[2]</sup>。对具有混沌特征的序列,其功率谱具有连续性、噪声背景和宽峰特征;对于周期性的确定性系统运动,其功率谱是离散的,仅包括基频及其谐波或分频;而对准周期运动,包含有各式各样的频率,其功率

谱也是分离的,但谱线并不像周期运动那样以某间隔的频率分离。根据这些特征,可以将周期运动、准周期运动和白噪声式的随机运动与混沌运动区分开。但在实际分析中,由于噪声的影响,以及对于周期很长但数据长度有限的序列,很难从谱特征上区分其运动模式<sup>[3]</sup>。

表 1 三江平原建三江分局地下水埋深观测资料基本情况

观测井名称	资料长度/d	观测频率	观测井类型	起止时间
创业农场 2 队长观井	706	5 d 一测	闲置水田井	1996-07-25 - 2006-06-30
创业农场 10 队长观井	703	5 d 一测	生活井	1996-07-25 - 2006-06-15
创业农场 17 队长观井	706	5 d 一测	生活井	1996-07-25 - 2006-06-30
洪河农场 601 号长观井	710	5 d 一测	生活手压井(水田区)	1996-07-20 - 2006-07-15

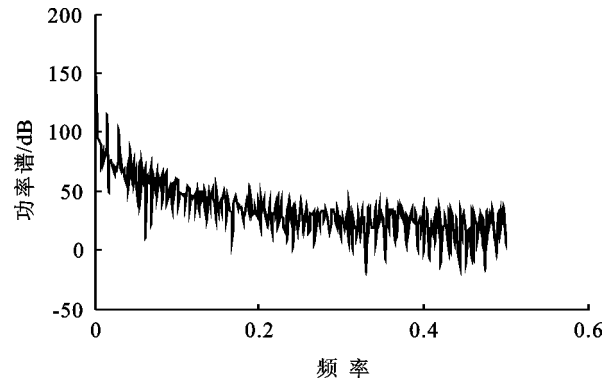


图 1 地下水埋深序列功率谱

创业农场和洪河农场地下水埋深序列功率谱形式如图 1。可以看出 4 个农场地下水埋深序列的功率谱具有连续性、噪声背景和宽峰特征,可能具有混沌特性。但由于功率谱分析是一种定性的分析方法,还需要进一步分析研究。

3 重构相空间

3.1 延迟时间的确定

自相关函数法<sup>[4]</sup>是非常成熟的独立求取时间延迟的方法,它主要是提取序列的线性相关性。一般地,对于一个混沌时间序列,可以先写出其自相关函数,然后做出自相关函数关于时间的函数图像。当自相关函数随延迟时间衰减明显时,取自相关函数首次过零时的时间作为延迟时间;当自相关函数随延迟时间衰减不明显时,取自相关函数下降到初始值时的  $(1 - 1/e)$ ,所对应的延迟时间能保证嵌入坐标的相关性较小,得到的时间就是重构相空间的延迟时间<sup>[5]</sup>。一般计算公式为

$$r = \frac{\sum_{t=1}^n (x_t - \bar{x})(x_{t+\tau} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (1)$$

利用自相关函数下降到初始值  $(1 - 1/e)$  时的时间作为延迟时间的选择,对三江平原地下水埋深的时间序列进行相关分析,得到各个农场相应的延迟时间,见表 2。

表 2 地下水埋深序列的延迟时间汇总

观测井点名称	创业农场 2 队	创业农场 10 队	创业农场 17 队	洪河农场 601 号
延迟时间	12	13	12	18

3.2 伪邻近点法确定嵌入维数

该法由 Kennel 等提出<sup>[6-7]</sup>:当嵌入维数较低时,由于轨

道未充分展开,相互挤压折叠,使得一些本来相距很远的相点折叠在一起(称作伪最邻近点);嵌入维数升高后,随着轨道的充分展开,挤压在一起的伪最邻近点逐渐分开,不再是“最近邻”。

地下水埋深系统重建相空间中,取前面自相关法所确定的延迟时间分别计算各个农场地下水埋深变化序列的伪最邻近点比率,最终求得的嵌入维数如表 3 所示。创业农场 2 队、10 队和 17 队的嵌入维数选取  $m = 7$ ,洪河农场 601 号长期观井的嵌入维数选取  $m = 9$  在理论上是比较合理的。

表 3 地下水埋深序列的嵌入维数汇总

观测井点名称	创业农场 2 队	创业农场 10 队	创业农场 17 队	洪河农场 601 号
嵌入维数 $m$	7	7	7 或 11	9

4 特征量的计算

4.1 Kolmogorov 熵

Kolmogorov 熵可以用来定义系统的不确定程度,是识别混沌序列的一个重要的特征量<sup>[8-9]</sup>。对于有序系统,  $K = 0$ ;对于随机系统,  $K = \infty$ ;  $0 < K < \infty$  便表征是混沌系统,且  $K$  越大,系统的混沌程度越严重。

通常用  $K_2$  作为熵  $K$  的一个估计,Grassberger-Procaccia 提出的计算式为

$$K_2 = \frac{1}{\ln 2} \ln \frac{C_m(r)}{C_{m+1}(r)} \quad (2)$$

式中:  $\tau$  ——延迟时间;  $C_m(r)$  ——嵌入相空间维数为  $m$  时的  $C(r)$  值;  $C_{m+1}(r)$  ——嵌入相空间维数为  $m + 1$  时的  $C(r)$  值。

各个农场地下水埋深序列的延迟时间取自表 2,则  $K_2$  在不同嵌入维数  $m + 1$  下的结果见表 4。随着嵌入维数的增加,三江平原各个农场地下水埋深序列的值趋于稳定,将  $K_2$  的稳定值作为熵的估计值。熵的值均大于零,这说明三江平原地下水埋深变化的实测序列存在混沌特性。但由于各个长观井所处地理位置不同,影响地下水埋深变化的因素也不相同,所以地下水埋深的运动也不相同,表现出有的地方周期性明显一些,所以熵的值就小,有的地方周期性不显著,所以熵值就大,这与熵的理论意义是相吻合的。

4.2 关联维

Packard 等建议用原始系统中某变量的延迟坐标来重构

相空间,Takens 证明了可以找到一个合适的嵌入维,即如果延迟坐标的维数  $m \geq 2d + 1$ ,  $d$  是动力系统的维数,在这个嵌入空间里可以把有规律的轨迹(吸引子)恢复出来,如果可以得到较小的分维数,那么就认为该动力系统具有混沌特征,该方法是一类最为常用的用以识别混沌特性的方法。

表 4 地下水埋深序列在不同嵌入维数  $m + 1$  下的  $K_2$  汇总

观测井点	延迟时间	嵌入维数				
		6	7	8	9	10
创业农场 2 队	12	0.087	0.051	0.040	0.036	0.038
创业农场 10 队	13	0.063	0.045	0.043	0.047	0.046
创业农场 17 队	12	0.095	0.070	0.064	0.061	0.064
洪河农场 601 号	18	0.013	0.014	0.015	0.015	0.015

但是,如果一个时间序列的时序性(即自相关性)显著时,在采用 G-P 算法计算其关联积分  $C(r, m)$  时,会误将相点在时序上的相关性当作一种状态空间几何<sup>[10]</sup>。为了防止出现这种情况,Theiler<sup>[11]</sup>提出通过设定 Theiler 窗口来解决这种问题,Grassberger<sup>[12]</sup>在提出关联维优化算法的同时,也建议对于存在时序相关性的序列要慷慨使用 Theiler 窗口参数,修正的关联维计算公式为

$$C(r, m) = \frac{2}{(N - w)(N + 1 - w)} \sum_{j=w}^{N-j} \sum_{i=1}^{N-j} (r - Y(t_i) - Y(t_j)) \quad (3)$$

式中: $Y(t_i)$ ——重构状态空间中的相点; $m$ ——状态空间嵌入维数; $N$ ——重建后状态空间数据点的数目,即  $N = n - (m - 1)$ ; $r$ ——以  $Y(t_i)$  为中心的  $m$  维嵌入空间中的球体半径,可视为标度尺度,即度量相点间距离的尺子长度; $\sum$ ——欧几里德范数; $w$ ——Theiler 窗口的大小,物理意义在于统计相空间中两点距离时,排除时间间隔小于  $w$  的点对,与延迟时间的意义类似,在这里取  $w = 0$ 。

各个农场地下水埋深序列的延迟时间取自表 2,则得到各序列的关联维计算结果如表 5。

表 5 地下水埋深序列的饱和关联维汇总

观测井点名称	创业农场 2 队	创业农场 10 队	创业农场 17 队	洪河农场 601 号
嵌入维数	8	12	9	10
关联维	2.106	2.537	2.569	2.574

从表 5 可以看出:各个农场地下水埋深序列的关联维均是较小的分维数,这表明三江平原地下水埋深变化的实测序列存在混沌特性。

5 基于关联度的局域加权线性回归预测

局域法假定预测中心点与其欧氏距离意义下的邻近点具有高度的相关性,但蒋传文等<sup>[13]</sup>研究表明,当重构相空间的嵌入维数较高时,以上假定并非总是成立,他们基于关联度对高嵌入维混沌时间序列进行了预测,却未将相点间相关性大小作用于预测模型。基于此,岳毅宏等<sup>[14]</sup>在局域线性回归预测法的基础上,提出一种基于关联度的局域加权线性回归预测法,该方法以关联度代替欧氏距离作为判别相点间相关性的准则,并将相点间相关性大小通过“加权”的方式作用于混沌序列预测模型,从而克服了前述局域线性回归预测

法的缺点。

5.1 基本理论与步骤

对于给定的混沌时间序列  $\{x_i\} (i = 1, 2, \dots, n)$ ,由相空间重构理论可以得到该序列在  $m$  维重构相空间中的相点为

$$X_t = [x_t, x_{t-1}, \dots, x_{t-(m-1)}],$$
$$t = (m - 1) + 1, (m - 2) + 2, \dots, n$$

为了便于叙述,将上式转化为

$$X_t = [x_t, x_{t-1}, \dots, x_{t-(m-1)}] = [x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(m)}],$$
$$t = (m - 1) + 1, (m - 2) + 2, \dots, n$$

若待预测的点为  $x_{n+1}$ ,则相空间中预测中心点(即预测的起始点)应为相点  $x_n$ 。此时,依次计算相空间中除  $x_n$  外的所有相点与  $x_n$  的关联度

$$r_{ni} = \frac{1}{m} \sum_{j=1}^m (\min + \max) / (x_n^{(j)} + \max),$$
$$i = (m - 1) + 1, (m - 2) + 2, \dots, n - 1$$

式中: $r_{ni}$ ——相点  $X_i$  与  $X_n$  的关联度; $0 \leq r_{ni} \leq 1$ ;——分辨系数,  $(0, 1)$ ;且

$$r_{ni}^{(j)} = |x_n^{(j)} - x_i^{(j)}|$$
$$\min = \min_{1 \leq j \leq m} \min_{(m-1)+1 \leq i \leq n-1} |x_n^{(j)} - x_i^{(j)}|$$
$$\max = \max_{1 \leq j \leq m} \max_{(m-1)+1 \leq i \leq n-1} |x_n^{(j)} - x_i^{(j)}|$$

此时,取阈值  $r$  为区间  $(0, 1)$  内的一个较大值,当  $r_{ni} \geq r$  时,则称相点  $X_i$  为中心点  $X_n$  的相关点。在经过计算后,假定  $X_n$  的所有相关点为  $X_{k_i} (i = 1, 2, \dots, p; (m - 1) + 1 \leq k_i \leq n - 1)$ ,定义相点  $X_{k_i}$  的权值为

$$w_i = r_{nk_i} / \sum_{j=1}^p r_{nk_j} \quad (4)$$

采用线性逼近,则式(4)可变换为

$$X_{k_i+1} = F^{(m)}(X_{k_i}) = A + B X_{k_i}, i = 1, 2, \dots, p \quad (5)$$

$$\text{式中: } A = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_m \end{bmatrix}; B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \dots & \dots & \dots & \dots \\ b_{m1} & b_{m2} & \dots & b_{mm} \end{bmatrix}。$$

于是得到预测模型:

$$x_{k_i+1} = a_1 + \sum_{j=1}^m b_{1j} x_{k_i-(j-1)}, i = 1, 2, \dots, p \quad (6)$$

关于预测模型式(6)的参数估计问题,王永忠等<sup>[15]</sup>在局域线性回归预测法采用了常用的最小二乘法,其缺陷在于将预测中心点  $X_n$  的所有相关点对预测的影响程度视为是相同的,但实际情况是中心点  $X_n$  的相关点对预测的影响程度是随着相关度的不同而变化的。相关度越大,影响就越大,反之亦然。因此,为了体现这种影响程度的不同,岳毅宏等<sup>[14]</sup>基于式(4)所定义的权值,采用加权最小二乘法对式(6)进行参数估计,得

$$\sum_{i=1}^p w_i (x_{k_i+1} - a_1 - \sum_{j=1}^m b_{1j} x_{k_i-(j-1)})^2 = \min \quad (7)$$

运用基于 Gauss-Newton 算法的最小二乘拟合及其 MATLAB 语言 nlinfit() 函数命令求出使得式(7)最小的参数  $a_1, b_{11}, b_{12}, \dots, b_{1m}$ ,将该参数代入预测模型公式(4),就完成了对混沌序列预测模型的建立。

预测时,首先以相点  $x_n$  作为预测中心点,得到点  $x_{n+1}$  的预测值  $\hat{x}_{n+1}$ ;然后在  $\hat{x}_{n+1}$  的基础上构造相点  $X_{n+1}$  并以  $X_{n+1}$  作为预测中心点,得到点  $X_{n+2}$  的预测值  $\hat{x}_{n+2}$ ;以此类推,可以得到测试序列中每个点的预测值。

5.2 应用预测

在进行预测时,首先 12 个数据作为测试数据;然后根据前面的计算结果确定出相空间重构的参数;最后,依照前述

的计算步骤进行预测拟合,预测精度见表 6。

从表 6 看出,在测试序列的预测中,相对误差绝对值小于 5 % 的点占总点数的 87.5 %,并且最大相对误差的绝对值是 8.73 %,可见预测效果是比较理想的。并且在预测开始时预测值与实际值的拟合精度很高、趋势比较一致,但随着预测步数的展开,预测精度和趋势开始变低,这源于混沌对初始值的敏感和噪声的影响,需要进一步改进。

表 6 地下水埋深预测精度汇总

观测井点	<i>m</i>	相对误差/ %												
		12	1	2	3	4	5	6	7	8	9	10	11	
创业 2 队	12	7	- 3.06	- 3.36	0	0.81	0	1.50	0.74	- 3.13	1.63	5.00	5.08	3.45
创业 10 队	13	7	2.07	- 0.29	- 0.54	- 0.20	0.91	2.83	1.81	2.07	0.46	1.63	1.12	- 5.57
创业 17 队	12	7	1.32	0.86	2.08	0	- 1.92	- 4.41	2.33	1.05	0.57	6.41	- 8.73	5.09
洪河 601 号	18	9	- 1.09	- 1.65	- 2.19	- 2.92	- 2.37	- 2.49	- 2.11	- 1.54	- 1.25	- 1.90	- 1.13	- 2.79

6 结 语

首先应用功率谱分析法分别对三江平原部分农场地下水埋深序列进行混沌定性分析,然后通过相空间重构理论,分别计算了农场地下水埋深序列的 Kolmogorov 熵、关联维,分析和计算表明:观测井点的地下水埋深序列存在明显的混沌特征,因而可以认为复杂的三江平原地下水埋深是混沌系统演化的结果。

从预测结果来看,预测的精度比较高,预测的趋势也非常准确,预测效果总的说来是理想的。在计算中,关联度阈值的选取是关键,取值将直接影响预测的效果。此外,数据的噪声、数据的长度等也影响预测的效果,这需要进一步的分析和研究并采取可能的措施,来提高预测精度。

对于不同性质的时间序列,应该采用不同的预测模型。由于三江平原地下水埋深序列中存在明显的混沌特征,那么在进行预测时,采用混沌预测方法得到的结果就会优于其他方法。只有对地下水埋深进行科学的预测,才能为三江平原的地下水资源开发规划、为三江平原的井灌水稻合理发展提供科学的决策依据。

参考文献:

[1] 付强. 农业水土资源系统分析与综合评价[M]. 北京: 中国水利水电出版社, 2005: 8-61, 254.

[2] 宋宇, 陈家军, 孙雄. 地下水水位时间序列中的混沌特征[J]. 水文地质工程地质, 2004(1): 14-18.

[3] 赵永龙. 水文动力系统混沌分析及其非线性预测[D]. 成都: 四川联合大学, 1997.

[4] 丁晶, 王文圣, 赵永龙. 长江日流量混沌变化特性研究: 相空间嵌入滞时的确定[J]. 水科学进展, 2003, 14(4): 407-411.

[5] 吕金虎, 陆君安, 陈士华. 混沌时间序列分析及其应用

[M]. 武汉: 武汉大学出版社, 2002: 49-59.

[6] Zaldivar J M, Gutierrez E. Forecasting high waters at Venice Lagoon using chaotic time series analysis and nonlinear neural networks[J]. Journal of Hydro. Informatics, 2000(2): 61-84.

[7] 丁晶, 王文圣, 赵永龙. 长江日流量混沌变化特性研究: 相空间嵌入维数的确定[J]. 水科学进展, 2003, 14(4): 412-416.

[8] 魏一鸣, 孙国栋, 胡平昭, 等. 九江年降水时间序列的混沌特性[J]. 江西科学, 1998, 6(3): 141-145.

[9] 李眉眉, 丁晶, 王文圣. 基于混沌理论的径流降尺度分析[J]. 四川大学学报: 工程科学版, 2004, 36(3): 14-19.

[10] 王文, 许武成. 对水文时间序列混沌特征参数估计问题的讨论[J]. 水科学进展, 2005, 16(4): 609-616.

[11] Theiler J. Spurious dimension from correlation algorithms applied to limited time-series data [J]. Phys. Rev. A., 1986, 34(3): 2427-2432.

[12] Grassberger P. An optimized box-assisted algorithm for fractal dimensions[J]. Phys. Lett. A., 1990, 148: 63-68.

[13] 蒋传文, 侯志俭, 张勇传. 基于关联度的高嵌入维混沌预测方法研究[J]. 系统工程与电子技术, 2002, 24(12): 65-66.

[14] 岳毅宏, 韩文秀, 张伟波. 基于关联度的混沌序列局域加权线性回归预测法[J]. 中国电机工程学报, 2004, 24(11): 17-20.

[15] 王永忠, 曾昭磐. 混沌时间序列的局域线性回归预测方法[J]. 厦门大学学报: 自然科学版, 1999, 38(4): 636-640.