

基于多变量相空间重构的地下水动态神经网络预测模型

刘卫林^{1,2}, 董增川¹, 陈南祥³, 董贵明¹

(1. 河海大学水文水资源与水利工程科学国家重点实验室, 江苏 南京 210098; 2. 南昌工程学院水利工程系, 江西 南昌 330028; 3. 华北水利水电学院岩土工程系, 河南 郑州 450008)

摘要: 从多元变量时间序列出发, 以相空间重构理论为基础, 结合偏最小二乘回归与神经网络方法, 提出了基于多变量相空间重构地下水动态预测的神经网络模型. 通过分别确定各个时间序列的时间延迟和嵌入维数的方法对地下水动态的多变量时间序列进行相空间重构, 采用偏最小二乘回归法提取对系统解释最强的成分作为网络输入, 利用神经网络模型进行各成分之间的非线性拟合. 将基于多变量相空间重构地下水动态预测的神经网络模型应用于黑龙洞泉域地下水位预测, 并通过分析与比较验证了该方法的有效性.

关键词: 地下水动态; 相空间重构; 神经网络; 偏最小二乘回归; 预测模型

中图分类号: TV211.1 文献标识码: A 文章编号: 1000-198X(2008)01-0014-05

地下水系统是一个受到多种因素影响的复杂的非线性动力系统^[1-2], 内部各因素之间以及内部因素与外部因素之间存在着十分复杂的交互作用和因果关系. 采用传统的预测方法如灰色系统方法、统计学方法等难以反映这种复杂的变化规律, 往往预测效果不是很理想. 近年来, 混沌理论和应用技术^[3-7]研究的不断深入以及非线性预测方法的发展, 为研究这一问题提供了新思路. 陈南祥等^[8-10]从地下水动态的单变量时间序列出发, 通过相空间重构技术建立了地下水位(或泉流量)预测模型, 并取得了较好的预测效果. 从理论上讲, 如果嵌入维数选取合理, 地下水系统水文要素的单变量时间序列也可取得较理想的预测效果. 但在实际应用中, 由于受数据容量和噪声等因素影响, 单变量时间序列重构的相空间往往不能十分准确地刻画原动力系统状态变量的演化轨迹, 预测质量受到一定程度的影响. 此外, 实际问题中也不能保证给定的单变量时间序列包含了重构动力系统的完整信息, 而多变量时间序列蕴含了更丰富更全面的系统信息, 从而能重构出更为准确的相空间, 提高预测质量^[11-12]. 基于此, 本文研究了地下水动态的多元变量时间序列预测方法, 提出了基于多变量相空间重构的地下水动态预测神经网络模型, 并用此模型对黑龙洞泉域的地下水位进行了预测.

1 基本原理

1.1 相空间重构

设地下水系统有 M 维多变量时间序列 $X_1, X_2, \dots, X_N$, 其中 $X_n = (x_{1,n}, x_{2,n}, \dots, x_{M,n})^T (n = 1, 2, \dots, N)$, 它对应系统 M 个连续变量 $X(t) = (x_1(t), x_2(t), \dots, x_M(t))^T$ 在时刻 t 的观测值, 即 $X_n = X(t_0 + n\Delta t)$, 其中 t_0 为初始时刻, Δt 为采样时间间隔. 类似单变量时间序列 ($M = 1$), 作多变量时间序列时间延迟重构

$$V_n = (x_{1,n}, x_{1,n-\tau_1}, \dots, x_{1,n-(m_1-1)\tau_1}, x_{2,n}, x_{2,n-\tau_2}, \dots, x_{2,n-(m_2-1)\tau_2}, \dots, x_{M,n}, x_{M,n-\tau_n}, \dots, x_{M,n-(m_M-1)\tau_M}) \quad (n = J_0, J_0 + 1, \dots, N)$$

(1)

式中 $J_0 = \max(m_i - 1)\tau_i + 1$, τ_i 和 m_i 分别为延迟时间和嵌入维数 ($i = 1, 2, \dots, M$). 根据 Takens 延迟嵌入原理, 只要 m_i 或 $m = \sum_{i=1}^M m_i$ 足够大, 则存在一个映射 $F: R^m \rightarrow R^m$ 使 $V_{n+1} = F(V_n)$. 这样, 上述 $N - J_0$ 个相点 V_n 连线描述了系统在 m 维相空间中随时间的演化过程, 从而在 m 维相空间中得到了一个相型. 从理论上

收稿日期: 2007-03-22
基金项目: 教育部科学技术研究重点项目(104197)
作者简介: 刘卫林(1978—), 男, 江西新余人, 博士研究生, 主要从事水资源系统分析与规划管理研究.

讲,有足够的信息把原系统的动态展现在多维相空间上.在相空间重构时,时间延迟 τ 和嵌入维数 m 的选取决定了重构效果的好坏并影响到后续的研究,而由于目前对多变量时间序列相空间重构研究比较少,重构参数的选取还处在探索阶段,本文采取分别求取各个时间序列时间延迟和嵌入维数的方法对地下水系统的多变量时间序列进行重构,对重构后的序列利用偏最小二乘回归法进行处理,提取对地下水动态解释最强的成分,避免了多变量时间序列相空间重构参数的复杂计算.单变量时间序列时间延迟 τ 和嵌入维数 m 选取的具体算法参见文献 [5-6,13].

1.2 偏最小二乘回归

偏最小二乘回归 (PLS) [14] 采用降维思想将系统多个变量进行综合与筛选并提取出少数对系统具有最佳解释能力的新的综合变量,这些综合变量相互独立.本文主要利用偏最小二乘法对重构后的序列提取对地下水动态解释最强的成分,以克服因独立选取各个时间序列相空间重构参数而引起重构序列之间的相关性.另外,通过偏最小二乘法对系统进行降维处理,以降低神经网络输入维数,简化网络结构,增强网络稳定性 [15]. 下面以单因变量偏最小二乘回归模型来说明该方法.

1.2.1 建模步骤

记地下水系统某一实测变量(因变量)为 $y \in \mathbf{R}^q$,由降雨量、蒸发、灌溉回归补给、侧向补给等影响因子组成的 p 个自变量为 $\{x_1, x_2, \dots, x_p\}$.为了研究因变量与自变量的统计关系,采集 q 个样本点,构成自变量与因变量的数据表 $X = (x_1, x_2, \dots, x_p)_{q \times p}$ 和 $Y = (y)_{q \times 1}$.偏最小二乘回归建模基本步骤可归纳如下:

a. 首先对数据进行标准化处理,记 E_0, F_0 分别为 X, Y 经标准化处理后的数据矩阵.

$$\begin{cases} E_{0i} = (x_i - E(x_i))/S_{x_i} & (i = 1, 2, \dots, p) \\ E_0 = (E_{01}, E_{02}, \dots, E_{0p})_{q \times p} & F_0 = (y - E(y))/S_y \end{cases} \quad (2)$$

式中: $E(x_i), E(y)$ —— x_i, y 的均值; S_{x_i}, S_y —— x_i, y 的均方差.

b. 分别从 E_0, F_0 中提取第 1 个主成分 $t_1 = E_0 w_1$ 和 $u_1 = F_0 c_1$,其中 w_1, c_1 分别是 E_0, F_0 的第 1 个轴,且 $\|w_1\| = 1, c_1 = 1, w_1 = E_0^T F_0 / \|E_0^T F_0\|^2$.分别作 E_0 和 F_0 对成分 t_1 的回归方程:

$$E_0 = E_1 + t_1 p_1^T \quad F_0 = F_1 + r_1 t_1 \quad (3)$$

式中: p_1, r_1 ——相应的回归系数向量, $p_1 = E_0^T t_1 / \|t_1\|^2, r_1 = F_0^T t_1 / \|t_1\|^2$; E_1, F_1 ——回归方程的残差.

c. 以 E_1, F_1 代替 E_0, F_0 用上述方法提取第 2 个主成分 $t_2 = E_1 w_2, u_2 = F_1 r_2$ 以及 E_0, F_0 的第 2 个主轴 w_2 ,其中 $w_2 = E_1^T F_0 / \|E_1^T F_0\|^2$,作 E_1, F_1 对 t_2 的回归方程:

$$E_2 = E_1 + t_2 p_2^T \quad F_2 = F_1 + r_2 t_2 \quad (4)$$

式中: $p_2 = E_1^T t_2 / \|t_2\|^2, r_2 = F_1^T t_2 / \|t_2\|^2$.

d. 第 h 步($h = 3, 4, \dots, A, A$ 为 X 的秩),推求第 h 个主成分 t_h 和 r_h 以及第 h 个主轴.

1.2.2 交叉有效性检验

一般不需要选用全部成分来建模,而是选用前 h 个成分($h < A$) $t_1, t_2, \dots, t_h$ 就可以得到预报性能良好的模型. h 的选取可采用类似于抽样测试法的工作方式:除去某个样本点 i 的所有样本点组成新的样本,使用 h 个成分拟合一个回归方程,得到 y_i 在样本点 i 上的拟合值 $\hat{y}_{(h-i)}$.对每一个样本点重复上述计算,并定义 y 的预测误差平方根和为 R_h ,有

$$R_h = \sum_{i=1}^n (y_i - \hat{y}_{(h-i)})^2 \quad (5)$$

显然,如果回归方程的稳健性不好,误差很大,它对样本点的变动就会十分敏感,这种扰动误差的作用就会加大 R_h 值,当 R_h 最小时,此时的 h 即为所求的成分数.

1.3 神经网络模型

基于相空间重构的时间序列预测的常用方法有全域法、局域法和基于神经网络的时间序列预测方法等,其中后者更适用于地下水动态的多变量时间序列预测的非线性问题 [15]. 本文采用 3 层 BP 网络建模. BP 网络模型是 ANN 模型中最典型、应用最广泛的一种多层前馈网络模型,它包含输入层、隐含层及输出层.设输入层有 d 个结点 $t_1, t_2, \dots, t_d$,输出层为 1 个结点 y ,隐含层数目由参考文献 [16] 的方法确定.输入层与隐含层之间的权为 V_{ij} ,阈值为 θ_j ,隐含层与输出层之间的权为 W_{j0} ,阈值为 γ ,则各层神经元的输出满足

$$y_0 = f(\sum_{j=1}^h W_{j0} x'_j - \gamma) \quad x'_j = f(\sum_{i=1}^d V_{ij} t_i - \theta_j) \tag{6}$$

输出误差定义为

$$E_{\text{总}} = \frac{1}{2} \sum_{k=1}^q (\hat{y}_0^k - y_0^k)^2 \tag{7}$$

式中： q ——样本数； $\hat{y}_0^k, y_0^k$ ——第 k 个样本学习算法的输出值和实际值。BP 模型通过正向传播和反向传播来调整神经元的权值，使得输出误差 $E_{\text{总}}$ 最小，具体算法详见文献 [16]。

2 模型的建立

相空间重构方法中最常用的是延迟坐标法，但用该方法重构时间序列其重构质量往往依赖于重构窗口的选择，由于受数据容量和噪声等因素影响，坐标间难免存在线性依赖；另一方面，神经网络虽然可以以其强非线性逼近能力通过输入输出样本的学习自动提取蕴含于系统中的复杂非线性映射规则，但是如果输入变量之间的相关性过强，就会影响网络的性能，破坏预测结果。而偏最小二乘回归方法可以很好地解决以上问题。偏最小二乘回归方法利用降维的思想将系统多个变量进行综合与筛选并提取出少数对系统具有最佳解释能力的新的综合变量，用这些互不相关的综合变量来描述多维空间的绝大部分动态特性。因此，可以利用偏最小二乘回归方法在延迟重构相空间的基础上对重构时间序列进行信息提取，选择对系统解释最佳的成分，消除坐标之间的线性依赖性，同时以这些提取的成分作为神经网络输入，解决网络输入相关性问题的。基于多变量相空间重构的地下水动态预测神经网络模型的具体建模步骤如下：

- a. 选取地下水动态变化的影响因素。影响地下水动态变化的因素主要包括降雨、蒸发、河流侧向补给等，通常不同因素序列的选取对预测结果影响不同，需根据具体问题和数据收集难易程度选取变量序列。
- b. 确定延滞时间 τ 和嵌入维数 m 。对于地下水动态的多变量时间序列的时间延迟 τ 和嵌入维数 m 的选取，可以针对各变量子序列 $\{x_{i,m}\}$ 采用单变量时间序列确定延迟时间和嵌入维数的方法分别选取 τ_i 和 m_i ，然后利用所得的参数按照式 (1) 对地下水动态的多变量时间序列进行相空间重构得到 $N - J_0$ 相点 V_n 。
- c. 利用偏最小二乘回归法提取主成分。对地下水系统多变量时间序列重构得到的每个相点 V_n ，首先按式 (2) 对 V_n 进行标准化处理，然后利用式 (3) (4) 进行主成分提取，并采用交叉有效性利用式 (5) 计算 R_h ，根据 R_h 值变化来确定提取的主成分数。一般以 R_h 达到最小值所对应的 h 作为提取的主成分数，得到互不相关的 h 个成分 $t_1, t_2, \dots, t_h$ 构成主成分转换矩阵 T 。
- d. 网络结构的建立。根据偏最小二乘回归得到的 h 个主成分，以 h 作为网络的输入层节点数，输出层节点为 1，隐含层节点数通过试错法确定。从重构的地下水多变量时间序列中提取学习样本，经主成分转换矩阵 T 转换后构成训练集，对网络进行训练，直至达到误差精度要求为止。
- e. 预测。网络训练成功后，以此时的网络参数进行预测，将最后 m 维相空间相点 V_N 通过主成分转换矩阵 T 提取 h 个主成分加入到网络输入中，对地下水位（或泉流量） $x_{1,N+l}$ 进行预测。

3 应用实例

以黑龙洞泉域 1958 ~ 2002 年地下水动态数据序列为依据，考虑到资料获取的难易程度，选取降雨量作为影响因素，构成地下水动态的多变量时间序列。地下水位动态曲线与降雨量的关系如图 1 所示。以 1958 ~ 1997 年地下水位及降雨量数据作为相空间重构和神经网络学习的基本数据，将 1998 ~ 2002 年的实测数据用于预测检验。

3.1 确定相空间饱和嵌入维数 m 和延迟时间 τ

采用互信息法分别计算地下水位以及降雨量时间序列的互信息函数，并取其第 1 个极小值点作为延迟时间，求得地下水水位时间序列的延迟时间 $\tau_1 = 2$ ，降雨量时间序列的延迟时间 $\tau_2 = 3$ ，嵌入维数的计算采用

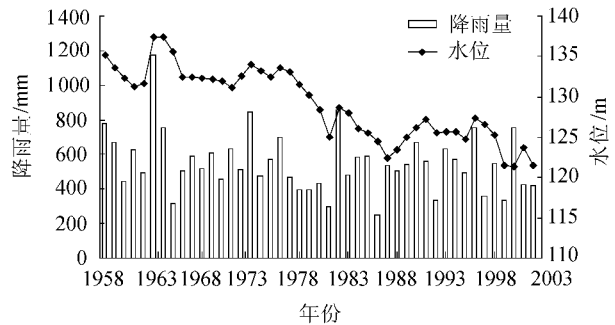


图 1 地下水位动态曲线与降雨量的关系

Fig.1 Relationship between groundwater table and precipitation

文献 [17] 提供的方法,计算得到地下水水位时间序列的嵌入维数 $m_1 = 2$,降雨量时间序列的嵌入维数 $m_2 = 8$. 利用所得的参数按照式 (1) 重构相空间.

3.2 用 PLS 提取主成分

对重构序列按式 (2) 进行标准化处理,然后根据偏最小二乘回归方法提取成分 T ,并采用交叉有效性确定主成分个数, R_h 随成分个数变化情况如图 2 所示. 从图 2 可以看出,当主成分数为 8 时, R_h 值达到最小,因此确定主成分个数为 8. 这 8 个主成分 $t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8$ 的表达式为

$$\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \\ t_8 \end{bmatrix} = \begin{bmatrix} 0.2064 & -0.4605 & -0.4176 & -0.6502 & -0.3734 & -0.1918 & 0.2104 & -0.4422 \\ 0.4511 & -0.0046 & 0.1995 & -0.0224 & -0.2815 & -0.4833 & 0.2221 & 0.4759 \\ 0.1440 & -0.2765 & 0.0921 & 0.2873 & 0.8758 & 0.5717 & 0.0805 & -0.1017 \\ 0.3907 & 0.0353 & -0.1321 & 0.0428 & 0.2170 & 0.4117 & 0.0117 & 0.5925 \\ -0.5054 & -0.5146 & -0.3999 & -0.6145 & -0.1187 & -0.4321 & -0.3260 & 0.3287 \\ 0.2170 & 0.2196 & 0.1964 & -0.4763 & -0.3632 & -0.4020 & -0.7140 & -0.2904 \\ -0.1335 & -0.4552 & 0.3363 & 0.6331 & 0.1493 & -0.2175 & -0.3823 & -0.1725 \\ 0.2447 & 0.3744 & 0.7194 & 0.3353 & 0.3530 & 0.4730 & 0.0797 & 0.0160 \\ -0.4386 & -0.3082 & 0.2582 & -0.1252 & -0.0697 & 0.3943 & 0.5363 & 0.2989 \\ -0.0884 & -0.4675 & -0.4580 & -0.4707 & -0.8148 & 0.0710 & -0.3294 & -0.0281 \end{bmatrix}^T \begin{bmatrix} x_{1,n-2} \\ x_{1,n} \\ x_{2,n} \\ x_{2,n-3} \\ x_{2,n-6} \\ x_{2,n-9} \\ x_{2,n-12} \\ x_{2,n-15} \\ x_{2,n-18} \\ x_{2,n-21} \end{bmatrix}$$

3.3 建立网络

将提取的 8 个主成分 $t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8$ 作为神经网络的输入,建立 8 个输入节点、9 个隐含层节点、1 个输出节点的神经网络模型,经 62 次迭代,收敛误差符合精度要求,学习结束. 网络学习的拟合值与实测值的比较如图 3 所示.

3.4 预测

根据训练好的网络,预测 1998~2002 年的地下水位,并以实测数据进行对比(图 3). 为验证方法的有效性,对地下水水位建立单变量时间序列相空间重构神经网络模型,在确定预测模型的结构时,直接分别取 $m = 2, 3, 4, \dots$ 维来重构相空间,以最好的模拟预测结果来筛选嵌入维数. 经反复试算, $\tau = 2, m = 4$ 时误差最小,此时网络参数为输入节点数为 4,隐含层节点数为 10,输出层节点数为 1. 预测结果见表 1. 由表 1 可以看出,2 种模型的计算值和实际值都比较接近,但多变量时间序列模型的平均相对误差为 0.25%,而单变量模型为 0.33%. 因此,总体上多变量时间序列模型较单变量模型预测效果要好些.

4 结 论

本文从地下水系统多元变量时间序列出发,以相空间重构理论为基础,将偏最小二乘回归与神经网络结合的方法引入到地下水动态的多变量时间序列建模与预测中,提出了基于多变量相空间重构地下水动态预测的神经网络模型,对地下水动态多变量时间序列

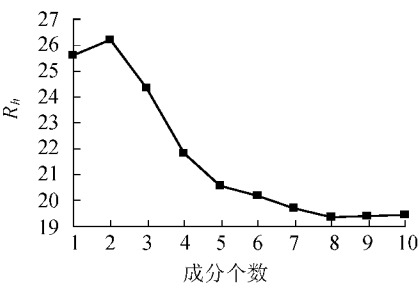


图 2 R_h 随成分个数变化的曲线
Fig.2 Variation of R_h with number of principal components

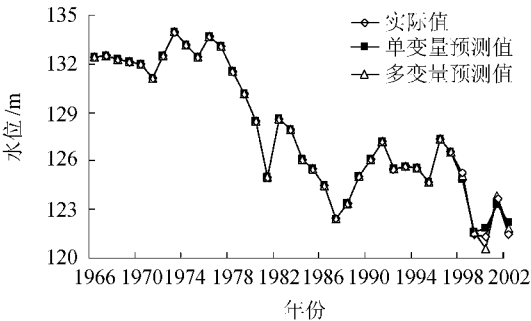


图 3 模型预测值与实际值的比较
Fig.3 Comparison of predicted result of model with observed data

表 1 黑龙洞泉域地下水位预测结果
Table 1 Predicted result of groundwater table for Heilongdong spring area

年份	实际值/m	多变量时间序列		单变量时间序列	
		预测值/m	相对误差/%	预测值/m	相对误差/%
1998	125.25	125.11	-0.11	124.86	-0.31
1999	121.50	121.61	0.09	121.60	0.08
2000	121.33	120.58	-0.62	121.83	0.41
2001	123.66	123.81	0.12	123.32	-0.27
2002	121.49	121.86	0.30	122.17	0.56

预测进行了有益的尝试.黑龙洞泉域地下水动态预测结果表明,基于多变量相空间重构地下水动态预测的神经网络模型,概念清晰,计算简便,具有较好的应用价值,为地下水动态预测提供了一种新思路,也可为多变量影响下的其他复杂系统预测提供借鉴.

由于资料限制,本文仅考虑了降雨量,而在实际问题中,地下水动态系统往往受降雨、蒸发、灌溉回归补给、侧向补给等因素影响,因此,对于资料丰富的地区,有必要对影响地下水动态变化的主要因素进行建模,以便更好地揭示地下水动态变化特征.

参考文献:

- [1] 秦四清,张倬元,王仕天.地下水动态预测的非线性动力学方法[J].工程勘察,1995(2):38-41.
- [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] PACKARD N H, CRUTCHFIELD J P, FARMER J D, et al. Geometry from a time series[J]. Physical Review Letter, 1980, 45: 712-716.
- [4] TAKENS F. Detecting strange attractors in turbulence[C]//RAND D A, YOUNG L S. Dynamical Systems and Turbulence. Berlin: Springer-Verlag, 1981: 366-381.
- [5] GRASSBERGER P, PROCACCIA I. Measuring the strangeness of strange attractor[J]. Physica D, 1983, 9: 189-208.
- [6] KIM H S, EYKHOLT R, SALAS J D. Nonlinear dynamics, delay times and embedding window[J]. Physica D, 1999, 127: 48-60.
- [7] WOLF A, SWIFT J B, SWINNEY H L, et al. Determining lyapunov exponents from a time series[J]. Physica D, 1985, 16: 285-317.
- [8] 陈南祥,曹连海,徐建新.地下水预报的相空间重构神经网络模型研究[J].西北农林科技大学学报:自然科学版,2005,33(12):139-142.
- [9] 宋宇,陈家军,孙雄.基于小波函数的地下水水位预测[J].工程勘察,2006(4):20-22.
- [10] 李贵明.基于相空间重构与神经网络耦合的岩溶泉流量预测模型[J].工程勘察,2006(4):30-32.
- [11] CAO Liang-yue, MEES A, JUDD K. Dynamics from multivariate time series[J]. Physica D, 1998, 121: 75-88.
- [12] 王海燕,盛田瀚,张进.多变量时间序列复杂系统的相空间重构[J].东南大学学报:自然科学版,2003,33(1):115-118.
- [13] 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.
- [14] 王惠文.偏最小二乘回归方法及其应用[M].北京:国防工业出版社,1999.
- [15] 陈南祥,黄强,曹连海.基于偏最小二乘回归与神经网络耦合的岩溶泉预报模型[J].水利学报,2004(9):68-72.
- [16] 张立明.人工神经网络的模型及其应用[M].上海:复旦大学出版社,1993.
- [17] CAO Liang-yue. Practical method for determining the minimum embedding dimension of a scalar time series[J]. Physica D, 1997, 110: 43-50.

Neural network model for groundwater prediction based on multivariate phase space reconstruction

LIU Wei-lin^{1,2}, DONG Zeng-chuan¹, CHEN Nan-xiang³, DONG Gui-ming¹

(1. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China;

2. Department of Hydraulic Engineering, Nanchang Institute of Technology, Nanchang 330028, China;

3. Department of Geotechnical Engineering, North China Institute of Water Conservancy and Hydroelectric Power, Zhengzhou 450008, China)

Abstract: Based on the phase space reconstruction theory for multivariate time series, a neural network model for groundwater prediction was put forward by combination with the neural network and partial least square regression methods. With the model, the phase space reconstruction for multivariate time series of groundwater was performed by determination of the time delay and embedded dimensions of each time series. By use of the partial least square method, the components which could reflect the characteristics of the system to the greatest extent were extracted and taken as the input of the network, and the neural network model was used for nonlinear fitting of each component. The validity of the model was verified by its application to groundwater prediction for Heilongdong spring area.

Key words: dynamic state of groundwater; phase space reconstruction; neural network; partial least square regression; prediction model