

马尔科夫链与 GM (1, 1) 模型在岩溶地下水 四氯化碳污染预测中的应用对比

刘喜坤¹, 张双圣^{1, 2}, 卜庆生³, 于向辉¹

(1. 徐州市城区水资源管理处, 江苏 徐州 221018; 2. 中国矿业大学环境与测绘学院, 江苏 徐州 221116;
3. 徐州市水利工程建设监理中心, 江苏 徐州 221018)

摘要: 为选择更优的岩溶地下水水质预测模型, 以七里沟岩溶地下水源地四氯化碳污染区为研究区域, 选用马尔科夫链与 GM (1, 1) 模型分别对四氯化碳污染进行预测研究, 并将预测值与实测值进行对比分析, 探讨两种预测模型的适用性及差异性。结果表明: 研究区四氯化碳污染监测数据具有一定的波动性, 马尔科夫链模型预测结果拟合程度更好, GM (1, 1) 适用于具有一定变化趋势的水质监测数据。

关键词: 马尔科夫链; GM (1, 1); 岩溶地下水; 四氯化碳; 预测

中图分类号: X52

文献标识码: B

文章编号: 1007-7839 (2017) 07-0001-05

Comparative analysis of forecasts in karst groundwater pollution based on Markov Chain and GM (1, 1)

LIU Xikun¹, ZHANG Shuangsheng^{1, 2}, BU Qingsheng³, YU Xianghui¹

(1. Xuzhou City Water Resource Administrative Office, Xuzhou 221018, Jiangsu; 2. School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, Jiangsu; 3. Xuzhou Water Project Construction Supervision Center, Xuzhou 221018, Jiangsu)

Abstract: In order to choose a better prediction model for the karst underground water quality, Markov chain and GM (1, 1) were taken separately to predict the level of CCl₄ in contaminated area of Qiligou water source. And the applicability and difference of the two prediction models were discussed through the comparison and analysis between measured values and predicted values. The results indicated that Markov chain prediction model showed higher fitting for the measured values with regular fluctuation, and GM (1, 1) was suitable to the water quality monitoring data with certain tendency of change.

Key words: Markov chain; GM (1, 1); Karst underground water; CCl₄; prediction

0 引言

水质预测是水环境规划、评价和管理工作的基础, 通过预测, 可以了解当地水域环境质量演变趋

势, 进而及时发现水质恶化的原因并采取相应的治理措施。《全国地下水污染防治规划(2011-2020年)》强调建立地下水污染风险防范体系, 构建地下水污染预报预测系统。

收稿日期: 2017-04-14

基金项目: 江苏省水利科技项目(2016056); 水体污染控制与治理科技重大专项(2015ZX07406005)

作者简介: 刘喜坤(1973-), 男, 博士, 教授级高级工程师, 主要从事水资源管理与水污染控制研究工作。

目前,地下水水质预测方法可分为两类,确定性方法和不确定性方法^[1]。其中确定性方法主要包括数值模拟法、解析法等;不确定性方法包括数理统计预测法、灰色系统理论预测法、神经网络模型预测法等^[2-3]。受水文、水质监测条件的限制,水质预测过程中往往缺乏长期的水质资料,导致确定性方法预测结果不够理想,尤其是岩溶地下水高度的非均质性,进一步加重了数值模拟的不确定性,因此不确定性方法在水质预测中得到广泛应用。

本文以徐州七里沟岩溶地下水水源地四氯化碳污染区为研究区域,分别运用马尔科夫链与GM(1,1)模型对四氯化碳进行污染预测研究,并对预测结果进行分析,探讨两种预测模型的适用性及差异性,以期对岩溶地下水水质预测提供借鉴和参考。

1 模型基本原理

1.1 马尔科夫链模型

马尔科夫分析方法主要用于分析随机事件未来发展变化的趋势。马尔科夫过程是具有无后效性的随机过程,即当过程在 t_m 时刻所处的状态已知时,过程在 t_{m+1} 所处的状态只与 t_m 时刻状态有关,而与 t_m 时刻之前的状态无关。通常将时间与状态都离散的马尔科夫过程称为马尔科夫链^[4-5]。

设系统的状态有 n 个,系统在 t_m 时刻处于状态 i ,在下一时间 t_{m+1} 时转为状态 j 的概率为 p_{ij} ,则称 p_{ij} 为一步转移概率,其矩阵形成:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

若系统经过 k 次转移后在时刻 t_{m+k} 处于状态 j 的概率为 $p_{ij}^{(k)}$,则 $p_{ij}^{(k)}$ 称为 k 步转移概率。相应的构成 k 步转移概率矩阵 $P^{(k)}$,且 $P^{(k)} = P^k$ 。

如果起始状态概率向量为 $P(0)$,经过 k 步以后,过程处于 j 状态的概率由下式得出:

$$P(k) = P(0) \cdot P^{(k)}$$

转移矩阵的确定是既是马尔科夫过程求解中的一个难题,也是求解关键。学者对此进行了深入探讨,但由于数据的随机性,对其求解问题一直没有完全解决。

本文将 i 年 j 监测井四氯化碳的浓度 C_{ij} 视为该年份状态向量的分向量,则历年的状态向量矩阵可以表示为:

$$C = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{m1} & C_{m2} & \cdots & C_{mn} \end{bmatrix}$$

式中:

m —监测年数;

n —典型监测井的个数。

本文运用统计学原理,利用最小二乘法估计马尔科夫转移矩阵,一步转移矩阵计算公式如下:

$$P = (X^T X)^{-1} (X^T) Y$$

式中:

$$X = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{(m-1)1} & C_{(m-1)2} & \cdots & C_{(m-1)n} \end{bmatrix};$$

$$Y = \begin{bmatrix} C_{21} & C_{22} & \cdots & C_{2n} \\ C_{31} & C_{32} & \cdots & C_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{m1} & C_{m2} & \cdots & C_{mn} \end{bmatrix}$$

1.2 GM(1,1)模型

GM(1,1)模型是一阶微分方程模型^[6],其形式为:

$$\frac{dx}{dt} + ax = b \quad (1)$$

其中 x 表示变量。

设非负原始序列 $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n))$ 对 $X^{(0)}$ 作一次累加,得到生成数列为 $X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n))$ 其中, $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$, $k=1, 2, \cdots, n$ 。

根据灰色系统理论, $X^{(1)}$ 的GM(1,1)模型白化形式的微分方程为:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (2)$$

其中 a, b 为待定参数, a 称为发展系数,反映预测的发展态势; b 称作灰作用量,反映数据的发展关系。

将(2)式离散化,得到GM(1,1)模型的定义式,如下:

$$x^{(0)}(k+1) = a[-\frac{1}{2}(x^{(1)}(k) + x^{(1)}(k+1))] + b \quad (3)$$

其中 $k=1, 2, L, n-1$ 。

根据最小二乘法可得待定参数 a, b 。并带入(2)式, 可得此一阶线性微分方程的通解, 并离散化为:

$$\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a} \right] e^{-ak} + \frac{b}{a} \quad (4)$$

还原到原始数据得:

$$\begin{aligned} \hat{x}^{(0)}(k+1) &= \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \\ &= (1 - e^a) \left[x^{(0)}(1) - \frac{b}{a} \right] e^{-ak} \end{aligned} \quad (5)$$

公式(4)和(5)为 GM (1, 1) 模型用于灰色预测的表达式。

2 应用实例对比

2.1 研究区概况

研究区属于贾汪复式向斜水文地质区中的七里沟裂隙岩溶水亚区的一部分, 位于徐州市的东南郊。含水岩组由奥陶系和寒武系组成, 核部被 20 ~ 40 m 的第四系覆盖, 水源地东部边界以断陷盆地的白垩系地层为隔水边界; 西部边界以南华系城门组岩层构成隔水边界; 南部以地下水分水岭为零通量边界; 北部受张扭性废黄河断裂带影响, 形成条带状富水块段, 阻隔污染物北移。研究区总面积约 103 km²。

通过污染源调查判定, 位于研究区南部地下

水补给区的一农药厂为四氯化碳污染源。通过前期研究, 岩溶含水层污染区可分为三个亚区: 南部的污染源附近亚区、北部集中抽水排泄区和中间污染相对较轻的过渡区。受构造、岩溶发育及人工开采等因素影响, 研究区地氯化碳污染羽呈哑铃状分布^[7]。

为充分体现研究区四氯化碳污染状态, 选取位于研究区南部、中部及北部的 X-47、X-59 及 X-56 水井作为典型监测井, 进行四氯化碳污染预测, 见图 1。

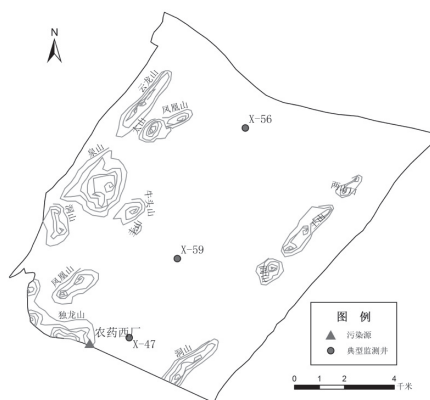


图 1 研究区地下水四氯化碳监测点位置图

2.2 对比分析

分别采用马尔科夫链与 GM (1, 1) 两种模型, 利用 MATLAB 软件编辑程序, 对 X-47、X-59 及 X-56 三眼典型监测井 2005 ~ 2015 年地下水四氯化碳浓度值进行模拟, 并以相对误差和标准差评价模型精度, 结果见表 1 ~ 表 3。

表 1 X-47 监测井四氯化碳实测年值与预测值对比

年份	监测值	预测值		相对误差		相对误差的标准偏差	
		马尔科夫链	GM (1, 1)	马尔科夫链	GM (1, 1)	马尔科夫链	GM (1, 1)
2005	91.18	91.18	91.18	0	0		
2006	50.55	57.83	55.86	-0.14	-0.10		
2007	55.73	36.59	48.04	0.34	0.14		
2008	58.69	35.02	41.32	0.40	0.30		
2009	17.32	23.83	35.53	-0.37	-1.05		
2010	21.20	25.66	30.56	-0.21	-0.44	0.28	0.44
2011	25.04	18.15	26.28	0.27	-0.05		
2012	30.70	20.37	22.60	0.34	0.26		
2013	13.60	14.90	19.44	-0.09	-0.43		
2014	20.40	16.88	16.72	0.17	0.18		
2015	19.36	12.71	14.38	0.34	0.26		

表 2 X-59 监测井四氯化碳实测年值与预测值对比

年份	监测值	预测值		相对误差		相对误差的标准偏差	
		马尔科夫链	GM (1, 1)	马尔科夫链	GM (1, 1)	马尔科夫链	GM (1, 1)
2005	31.40	31.40	31.40	0	0		
2006	17.41	20.71	17.76	-0.19	-0.02		
2007	19.30	15.65	17.49	0.19	0.09		
2008	12.40	13.80	17.23	-0.11	-0.39		
2009	11.51	12.40	16.96	-0.08	-0.47		
2010	23.45	11.38	16.71	0.51	0.29	0.28	0.29
2011	19.43	10.50	16.45	0.46	0.15		
2012	18.39	9.73	16.20	0.47	0.12		
2013	20.32	9.05	15.96	0.55	0.21		
2014	11.72	8.43	15.72	0.28	-0.34		
2015	12.01	7.86	15.48	0.35	-0.29		

表 3 X-56 监测井四氯化碳实测年值与预测值对比

年份	监测值	预测值		相对误差		相对误差的标准偏差	
		马尔科夫链	GM (1, 1)	马尔科夫链	GM (1, 1)	马尔科夫链	GM (1, 1)
2005	48.17	48.17	48.17	0	0		
2006	26.21	30.44	23.44	-0.16	0.11		
2007	17.21	16.58	23.06	0.04	-0.34		
2008	16.80	19.28	22.69	-0.15	-0.35		
2009	11.85	13.30	22.32	-0.12	-0.88		
2010	38.51	16.19	21.96	0.58	0.43	0.32	0.47
2011	27.94	11.71	21.61	0.58	0.23		
2012	28.09	13.94	21.26	0.50	0.24		
2013	24.80	10.37	20.92	0.58	0.16		
2014	14.69	12.05	20.58	0.18	-0.40		
2015	11.92	9.18	20.25	0.23	-0.70		

由表 1 ~ 表 3 可以看出, 两种预测模型的平均相对误差均小于 35%, 表明两种模型均具有一定的可信度。与 GM (1, 1) 模型相比, 马尔科夫链预测模型的相对误差的标准偏差更小, 表明马尔科夫链预测模型更适合于研究区地下水四氯化碳浓度的预测。

为直观表现两种预测模型的预测效果, 将监

测值与模拟值进行拟合, 见图 2。三眼典型监测井的监测数据均具有一定的波动性, 主要由于地下水四氯化碳浓度受构造、开采量、降水等因素的影响。其中马尔科夫链模型表现出一定的波动适应性, 而且比较适应于波动较小的数据的预测, 而 GM (1, 1) 则不能体现原始数据的波动性, 仅能把握数据的大体趋势。

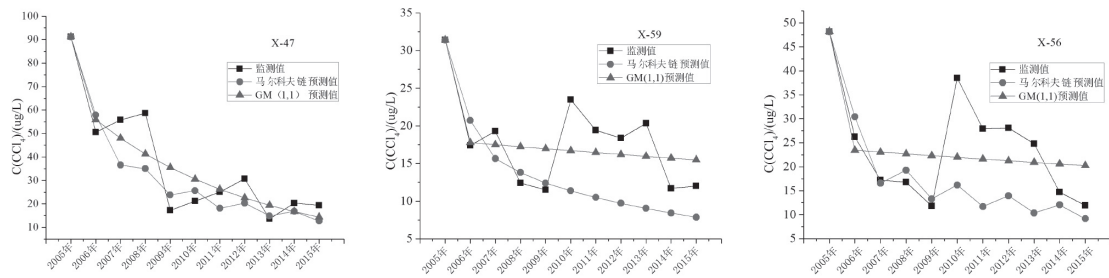


图2 监测值与模拟值拟合图

2.3 预测结果

通过 2.2 节分析可知, 马尔科夫链模型预测结果标准差较小, 而且研究区地下水四氯化碳浓度受开采量情况影响较大, 表现出一定的数据波动性, 因此采用马尔科夫链模型进行 2017 ~ 2020 年研究区地下水四氯化碳浓度的预测, 结果见图 3。

由图 3 可知, 2017 ~ 2020 年研究区地下水四氯化碳浓度呈下降趋势, 南北两端地下水四氯化碳浓度仍然高于中间区域, 而且靠近污染源的南部区域浓度高于北区。与 2017 年相比, 2020 年 X-47、X-59、X-56 三眼典型监测井四氯化碳浓度分别降低 16.02%、18.40%、12.76%。

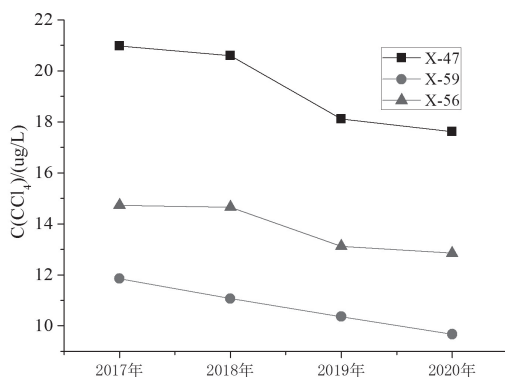


图3 马尔科夫链预测模型预测结果

3 结语

在一定可信度条件下, 马尔科夫链与 GM(1, 1) 均可用于地下水水质预测。对于具有一定波动性的水质监测数据, 马尔科夫链模型预测结果更理想, GM(1, 1) 适用于具有一定变化趋势的水质监测数据。对于受地下水开采量、降水等因素影响, 表现出一定波动性的研究区地下水四氯化碳污染预测, 首选马尔科夫链预测模型。

参考文献:

- [1] 杨海东, 肖宜, 王卓民, 等. 突发性水污染时间溯源方法[J]. 水科学进展, 2014, 25(1):122-129.
- [2] 冯江浪. 改进灰色马尔科夫模型及其在水资源预测中的利用[J]. 物探化探计算技术, 2010, 32(1):109-112.
- [3] 王宇博, 梁秀娟, 乔雨, 等. 地下水位预测模型对比分析研究[J]. 节水灌溉, 2015(7):58-61.
- [4] 卢丹. 改进的灰色—马尔科夫模型在地下水水质预测中的应用研究[J]. 水利规划与设计, 2016(6):86-89.
- [5] 李名升, 任晓霞, 佟连军. 马尔科夫链在环境质量预测中的应用[J]. 环境工程, 2007, 25(6):78-80.
- [6] 刘洪, 孙国曦, 曹瑞祥. GM(1, 1) 动态模型在吴江市地下水水位预测中的应用[J]. 地质灾害与环境保护, 2008, 19(3):47-51.
- [7] 韩宝平, 王小英, 朱雪强, 等. 某市岩溶地下水四氯化碳污染特征研究[J]. 环境科学学报, 2004, 24(6):982-988.

(责任编辑: 华智睿)