Doi:10.3969/j.issn.1007-1903.2015.04.013

基于灰色系统与线性回归方法的水质预测

张 洁，杨 庆，赵 杰

（北京市水文地质工程地质大队，北京 100195）

摘 要：以北京市石景山区某地地下水监测点多年监测数据为例，进行适宜性预测方法的验证。其中，总硬度灰色预测值的平均相对误差为2.62%比线性预测低5.65个百分点；溶解性总固体线性预测值的平均相对误差为2.33%比灰色预测低0.61个百分点，与通过拟合度R^2值大小选取的最优预测方法一致，表明通过拟合度R^2值大小来选取合理的预测方法是一种便捷、合理的技术手段。

关键词：地下水水质预测；灰色系统；线性回归方法；最小二乘法；拟合度；适应性；

中图分类号：X824 文献标识码：A 文章编号：1007-1903（2015）04-0067-05

0 引言

地下水系统是自然环境的重要组成部分，但随着人类社会的发展和生产规模的扩大，地下水污染越来越严重。因此，有必要采取有效措施，保护地下水。地下水水质预测是在水资源规划的基础上对地下水的评估和管理，是水资源保护的基础，可以及时了解地下水质的变化趋势。当前，国内外地下水水质预测已进入了实用化阶段[1~3]。伴随着科学技术的不断更新，预测的新方法还在不断涌现，目前主要的预测方法有马尔可夫法、灰色预测、神经网络模型预测法和线性回归分析法。本文根据预测方法适应性的不同，通过对比、判别两种方法预测的结果，选择更优化的方法对数据进行预测，有效的避免了单一预测方法的局限性，预测精度更高[4~5]。

1 预测方法

1.1灰色预测

1982年时我国的教授邓聚龙首先提出了灰色系统理论，并且提出它可以用连续的灰色微分模型来对系统以后的发展变化进行观察分析，并且做出预测[6]。灰色预测能够依据已有的少量信息，通过运算和分析得到预测结果。灰色预测中最为常用的就是G M(1,1)模型[7]。

原始时间序列：$X^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$

预测第$n+1$期，第$n+2$期，… 的值：$x^{(0)}(n+1),x^{(0)}(n+2),\cdots$

设相应的预测模型模拟序列为：$\hat{X}^{(0)}=(\hat{x}^{(0)}(1),\hat{x}^{(0)}(2),\cdots,\hat{x}^{(0)}(n))$

设$X^{(1)}$为$X^{(0)}$的1-AGO（即一次累加）序列：

$$x^{(1)}(i)=\sum_{m=1}^{i}x^{(0)}(m),i=1,2,3\cdots,n)$$

即：
$$\begin{cases}x^{(1)}(1)=x^{(0)}(1)\\x^{(1)}(i)=x^{(0)}(i)+x^{(1)}(i-1),i=2,\cdots,n\end{cases}$$

利用$X^{(1)}$计算G M(1,1)模型参数a、μ。

令$\hat{a}=[a,u]^T$ 则有：$\hat{a}=(B^TB)^{-1}B^TY_n$

式中：
$$B=\begin{bmatrix}-\frac{1}{2}(x^{(1)}(1)+x^{(1)}(2)) & 1\\-\frac{1}{2}(x^{(1)}(2)+x^{(1)}(3)) & 1\\\cdots & \cdots\\-\frac{1}{2}(x^{(1)}(n-1)+x^{(1)}(n)) & 1\end{bmatrix}$$

$$Y_n=[x^{(0)}(2),x^{(0)}(3),\cdots,x^{(0)}(n)]^T$$

由此获得G M(1,1)模型如下：

$$\hat{x}^{(1)}(i+1)=(x^{(0)}(1)-\frac{u}{a})e^{-ai}+\frac{u}{a}$$

作者简介：张洁（1991- ），男，主要研究方向为地下水环境监测与评价。E-mail:735498097@qq.com

如果预测模型的精度满足要求，则可用来预测，获得的预测值如下：

$\hat{x}^{(0)}(1)=\hat{x}^{(1)}(1)$， $\hat{x}^{(0)}(2)=\hat{x}^{(1)}(2)-\hat{x}^{(1)}(1)$，⋯，

$\hat{x}^{(0)}(n+1)=\hat{x}^{(1)}(n+1)-\hat{x}^{(1)}(n)$

$\hat{x}^{(0)}(n+2)=\hat{x}^{(1)}(n+2)-\hat{x}^{(1)}(n+1)$，⋯

已经建立的模型要通过残差检验和后验差检验才能运用预测。

后验差c和小误差概率p是后验差检验的两个指标，设S_{l2}和S_{22}分别是原始数列和残差数列的方差，即

$$S_{12}=\frac{1}{n-1}\sum_{k=1}^{n}(x^{(0)}(k)-\bar{x}^{(0)})^2$$

$$S_{22}=\frac{1}{n-1}\sum_{k=2}^{n}(\varepsilon^{(0)}(k)-\bar{\varepsilon}^{(0)})^2$$

其中：

$$\bar{x}^{(0)}=\frac{1}{n}\sum_{k=1}^{1}x^{(0)}(k)\ ,\quad \bar{\varepsilon}^{(0)}=\frac{1}{n}\sum_{k=2}^{1}\varepsilon^{(0)}(k)$$

用下式计算 C 和 P：

$C=S_2/S_1$, $P=P\{0.6745S1>\left|e^{(0)}(k)-\bar{e}^{(0)}\right|\}$,根据表1来判定模型的精度。只有模型能够满足后验差检验要求才能认为模型是合格。

表1 灰色预测模型精度表

精度等级	p 值	c 值
好	$0.95\leqslant p$	$c\leqslant 0.35$
合格	$0.80\leqslant p<0.95$	$0.35<c\leqslant 0.50$
勉强合格	$0.70\leqslant p<0.80$	$0.50<c\leqslant 0.65$
不合格	$p<0.70$	$0.65<c$

灰色系统理论有许多优点：①计算简单，可检验。②数据要求低，不需要太多数据即可预测。③预测精度高。由于灰色系统理论中的G M模型是近似呈指数增长的模型，被预测对象目标值的灰度和变化递变规律决定着预测的精度，因此可以判断出灰色系统理论方法比较适用于数据序列呈近似指数增长的数据的中期短期预测[8~10]。

1.2 线性回归预测

线性回归分析方法广泛的应用于各个科学领域。该方法不仅可以把藏于原始数据中的重要信息显露出来，而且可以得到变量之间存在的关系，用数学方式表达出来，得到原始数据特征，运用概率相关知识对数据的分析，判别其有效性。其中一个变量在受到其他变量的影响时，把这个变量称之为因变量，记为Y，其他的变量称之为自变量，记为X，这时相关关系式可记作：

$$Y=f(x)+\varepsilon$$

其中$f(x)$为当$X=x$时，因变量Y的均值，即 $f(x)=E(Y \quad X=x)$ 称$f(x)$为Y对X的回归函数，ε是Y对$f(x)$的偏差，它是一个随机变量，为了研究方便我们假定$E(\varepsilon)=0$。回归函数既可以是一元的函数也可以是多元的函数，即

$$Y=f(x_1,x_2,\cdots,x_m)+\varepsilon$$

其中 $f(x_1,x_2,\cdots,x_m)=E(Y\mid X_1=x_1,X_2=x_2,\cdots,X_m=x_m)$

为m元回归函数，统称为多元回归函数。

若回归函数 $f(x_1,x_2,\cdots,x_m)$ 中，$m=1$且 $f(x_1,x_2,\cdots,x_m)$ 是线性函数，则称$f(x)$为是一元线性回归函数；$m>1$时 $f(x_1,x_2,\cdots,x_m)$ 是多元线性函数，则称其为多元线性的回归函数；若所得的回归函数 $f(x_1,x_2,\cdots,x_m)$ 是非线性的函数，我们称之为非线性回归函数。要对整个回归方程做 F 检验和对变量之间关系做t检验并要预测执行度为95%的预测区间[11]。

线性回归分析除了有灰色系统理论的优点外，还可以很直观，快速的分析出数据之间的关系，并且可以确切的得出各个因素之间的拟合程度与相关程度的高低，提高预测方程式的效果。线性回归分析方法就是找到数据之间的线性关系，通过回归方程对数据预测，所以适合预测原始数据近似呈线性分布的数据。

1.3 适宜方法的选择

灰色系统法和线性回归法均有其各自的适宜性，灰色系统理论适宜预测呈指数型分布的数据，线性回归方法适宜预测呈线性分布的数据。对原始数据做最小二乘法处理得到原始数据的拟合度，并根据拟合度把线性回归方法和灰色系统理论方法结合。地下水水质预测分为3个步骤。

第一步：预测方法选择

根据最小二乘法原理得到的指数或线性的拟合度R^2大小来选取适合的方法进行预测，若水质指标的指数趋势线拟合度R^2高于线性的，则用灰色系统理论预测；反之水质指标的线性趋势线拟合度R^2高于指数的，则选取线性回归对指标进行预测。

第二步：预测

根据所选取的适宜性方法对数据预测，并对预测结果做原理性误差分析。

第三步：预测结果分析

预测数据与实际值的对比，分析比较两种方法的适应性。

2 适宜性预测方法的验证

2.1 预测方法的选择

选取北京市石景山某地为例进行预测，地下水溶解性总固体、总硬度多年检测数据见表2，趋势线散点图见图1、图2，比较R^2的大小得出适宜性的预测方法如表3。

表2 地下水多年水质数据

序号	总硬度（mg/L）	溶解性总固体（mg/L）	时间（年）
1	313	599	1996
2	295	678	1997
3	338	794	1998
4	340	652	1999
5	368	787	2000
6	354	775	2001
7	409	883	2002
8	425	901	2003
9	445	946	2004
10	471	1004	2005
11	-	-	2006
12	518	1059	2007
13	566	1136	2008
14	611	1171	2009

注：“-”表示无数据

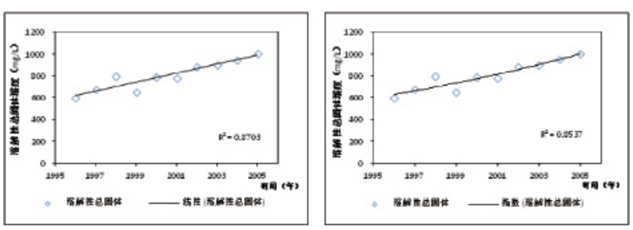

图1 溶解性总固体趋势线散点图

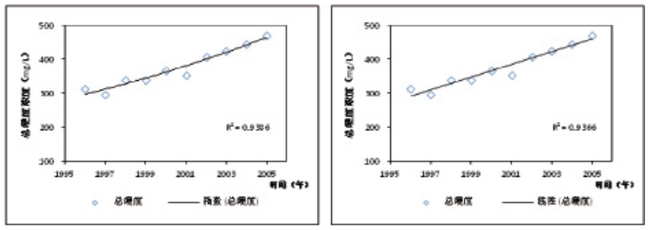

图2 总硬度趋势线散点图

表3 预测方法的选择

指标	指数 R^2	线性 R^2	预测方法
总硬度	0.939	0.937	灰色系统理论
溶解性总固体	0.854	0.870	线性回归分析

2.2 预测

（1）由于溶解性总固体的线性R^2大于指数R^2，应选择线性回归方法对溶解性总固体预测。用数理统计软件预测出来的结果见表4,此表给出了t检验的所有系数的检验结果以及标化/未标化的系数。通过系数表得到回归方程：

$$y = 40.830x - 80879.121$$

式中：x 是自变量年份，y 是因变量溶解性总固体。

误差检验：R^2=0.870，拟合效果比较好，线性成立。sig=0.000<0.005，t检验通过，模型成立。

（2）由于总硬度的指数拟合度R^2大于线性拟合度R^2，应选择灰色预测法预测总硬度。预测结果和误差分析见表5。平均相对误差为2.11%精度好，后检验误差C=0.15，P=1精度好，模型成立。

2.3 预测结果分析

用两种方法预测后3年的数据，与实测值对比见图3、图4。总硬度灰色预测值的平均相对误差为2.62%，比线性预测低5.65个百分点；溶解性总固体线性预测值的平均相对误差为2.33%，比线性预测低0.61个百分点，显然选择适宜性预测方法的预测值更准确。

3 结论

根据灰色预测和线性回归这两种典型预测方法适应性的不同，通过对原始数据做最小二乘法处理得到原始

表4 溶解性总固体线性预测结果

模型	回归方程系数	标准误差	t 检验	p 检验	95% 置信区间（上限）	95% 置信区间（下限）
总硬度浓度	-80879.121	11140.508	-7.26	0	-106569179	-55189.064
年份	40.83	5.569	7.332	0	27.988	53.672

表5 总硬度误差分析

原始数据	模拟值	残差	相对误差（%）
313	313	0.00	0.00
295	305	9.80	3.32
338	322	16.30	4.82
340	340	0.30	0.09
368	359	9.00	2.45
354	379	25.00	7.06
409	400	9.00	2.20
425	422	3.00	0.71
445	446	1.00	0.22
471	470	1.00	0.21

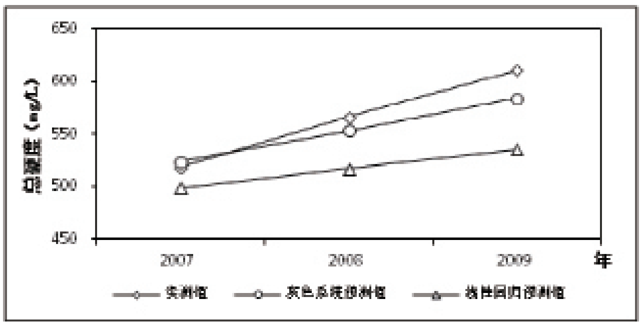

图3 总硬度预测值与实际值对比

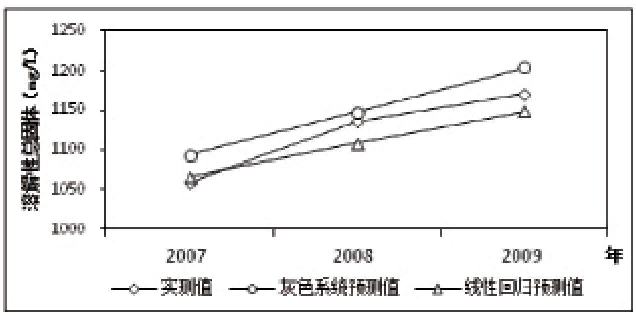

图4 溶解性总固体预测值与实际值对比

数据的拟合度，根据指数或线性的趋势线的拟合度R2大小来选取适合的方法进行预测，为以后地下水质预测及预警预报提供了一种新的思路。通过拟合度来选择所需的预测方法，不仅简单，而且预测的数值准确性更高。此种方法的不足是需要定点长期稳定的监测数据，未考虑水文地质、大气降水等外在条件，预测值只是趋势性预测。建议在有条件的情况下，可结合降水、水位等要素做多元回归分析，并根据社会环境、地质环境的阶段性变化做分段预测。

参考文献

[1]尤本胜，王明章，向速林. 地下水流量预测的线性神经网络法[J]. 贵州工业大学学报， 2004，33 (2): 60~621.

[2]于纪玉. 平原地区地下水环境预测方法研究[D]. 南京：河海大学，2003.

[3]梁文彪. 应用回归分析方法推求降水入渗补给系数[J]. 地下水，2002，24 (2): 72~731.

[4]赵延涛. 基于BP神经网络的地下水位预测[J]. 勘查科学技术，2001，4：7~10.

[5]张保祥，刘青勇，卢朝霞. 基于神经网络和遗传算法的济南市区岩溶地下水预报模型研究[J]. 山东农业大学学报（自然科学版），2004，35(3)：436~441.

[6]邓聚龙. 灰色系统基本方法[M]. 武汉：华中理工大学出版社，1987 .

[7]刘思峰，党耀国，方志耕. 灰色系统理论及其应用[M]. 北京:科学出版社, 2010.

[8]耿志民，李世军. 应用趋势分析灰色模块——均差插值法预报近期地下水情[J]. 水文地质工程地质，1994，21(1)：46~49.

[9]唐宗鑫，简文彬. 闽江下游水质预测的时间序列模型[J]. 水利科技，2002，(2)：7~9.

[10]卢崇飞，高惠璇，叶文虎. 环境数理统计应用及程序[M]. 北京：高等教育出版社，1988：33~38.

[11]陈志宏. 多元线性回归方法在地下水水位预测中的应用[J]. 北京地质，1999(3)：20~26.

Select the Suitability of Groundwater Quality Prediction Method

ZHANG Jie , YANG Qing , ZHAO Jie

(Hydrogeology and Engineering Geology Team of Beijing, Beijing 100195)

Abstract: To develop an efficient method, this paper selects the “Least Square Fit” method to fit the initial data, and selects the corresponding optimal prediction method by the value of R^2. To verify the feasibility of the method, both approaches have been tested by the data from a long-term groundwater monitoring points in Shijingshan District, Beijing. By applying the grey system method, the relative error for total hardness is 2.62%, which is 5.65% lower than the result of linear regression method. The relative error for TDS is 2.33% by grey system method, which is 0.61% lower than the result of linear regression method. The results are similar to that by the value of R^2. Namely, it is a rational and efficient technical method to select the suitable approach for groundwater quality prediction .

Keywords: Groundwater quality prediction; Grey system method; Linear regression method; Least Square Fit; Fitness ; Suitability

（上接第37页）

Forming Condition and Characteristics Analysis on the Qianhuoshiling Debris Flow Gully in Miyun County, Beijing

DING Xia

（Beijing Geological Engineering Design Institute, beijing 101500）

Abstract: There are about 856 debris flow gullies in Beijing area, located at Yanshan Mountains in the northern area of Beijing and Taihang Mountains in the western area of Beijing. Influenced by topography, geological conditions, rainfall conditions and human activities, the debris flow type is mainly rain gully debris flow. The Qianhuoshiling debris flow gully is a typical debris flow gully in the Yanshan Mountains. On the basis of geological survey and exploration, this paper comprehensively analyzes the forming conditions and characteristics of the Qianhuoshiling debris flow gully from geological conditions, topography, rainfall conditions and human activities, while the development of debris flow gully is analyzed in depth. According to the design criteria of once rainstorm in every10 years, the motion characteristic parameters of debris flow are calculated respectively in three sections of the upstream, midstream, headstream of the debris flow gully. We come to the conclusion that the debris flow belongs to the water stone flow and is in the rest period of the development stage at present. High intensity concentrated rainfall is the motivating factor for the debris flow. And peak flow landslides and other motion parameters gradually increase along the debris flow gully. The analysis of the article is of great significance to the selection of the control scheme of the debris flow gully. At the same time, it has a reference for the prevention and control of debris flow along the Yanshan Mountains.

Keywords: Debris flow; Characteristic analysis; Geology disaster