

文章编号:1009-3087(2013)增刊2-0055-06

使用秩次集对方法预测地下水位动态变化

何思为¹, 南卓铜^{2*}

(1. 四川大学 水利水电学院, 四川 成都 610065; 2. 中国科学院 寒区旱区环境与工程研究所, 甘肃 兰州 730000)

摘要:为了说明秩次集对分析方法在预测地下水位动态变化中的能力以及不同的历史集合容量和后续值个数对该方法预测结果的影响,结合甘肃省白银市景泰县2处地下水位测站的长序列观测资料,研究了不同的历史集合容量对单后续值秩次集对分析方法和多后续值秩次集对分析方法预测效果的影响。首先,取不同的历史集合容量 $T(3, 4, 5, 6, 8, 10, 12, 15)$,在单后续值的情况下分析了不同的历史集合容量对预测误差的影响。然后,分12种情形研究了在历史集合容量 T 固定的情形下不同的后续值个数 n 对预测精度的影响。研究发现,秩次集对分析方法在预测该地区地下水位动态变化时,若历史集合容量 T 小于8,误差基本保持稳定,但当 T 大于8,则误差会急剧增加;秩次集对方法在预测地下水位一个短时间序列内的变化精度较高,但是对地下水位动态变化长序列的模拟能力较差;秩次集对方法在预测时趋势误差累积相当严重;相对于模拟单一后续值,应用秩次集对分析方法预测 $n(n > 1)$ 个后续值可以较有效减缓预测时趋势误差的累积,但用于长序列预测精度仍然不高。

关键词:秩次集对分析;地下水;地下水位动态;预测

中图分类号:TV121.3

文献标志码:A

Application of Rank Set Pair Analysis Method to Predicting Groundwater Dynamics

HE Si-wei¹, NAN Zhuo-tong^{2*}

(1. College of Hydraulic and Hydroelectric Eng., Sichuan Univ., Chengdu 610065, China;

2. Cold and Arid Regions Environmental and Eng. Research Inst., CAS, Lanzhou 730000, China)

Abstract: In order to investigate the performance of rank set pair analysis (RSPA) in predicting groundwater dynamics and its response to the abovementioned two parameters, a series of scenarios were carried out by using the long-term groundwater table records in the Jingtai site of the Baiyin county, Gansu province. First, the size of historical set T was assigned as one of eight different values ranging 3 to 15 respectively. The effect of historical set size on prediction error was studied, in the case of single subsequent value. Then, by keeping historical set size T fixed, the influence of number of subsequent value n on prediction error was investigated. The study revealed that when historical set size was less than 8, the error was steady and acceptable and when the size became greater than 8, the error might remarkably increase. RSPA performed well in prediction of groundwater dynamics within a short term, however, poorly in predicting a long-term dynamics of groundwater. RSPA had serious disadvantage in error accumulation occurred in every step. In comparison with single subsequence value, the multi-subsequence value could make the accumulation of error of RSPA increases slowly, yet, it still could not fulfill the requirement of predicting a long-term groundwater dynamics.

Key words: rank set pair analysis (RSPA); groundwater; groundwater dynamics; prediction

地下水位动态变化预测在合理开采地下水^[1]、植被恢复^[2]、生态环境维护^[3]以及河流流量补给分析^[4-5]等生产和生活的诸多方面有重要的作用和意义。目前,地下水位预测常用的方法有回归分析

法^[6]、时间序列法^[7]、神经网络方法^[8]、小波变换方法^[9]等。由于地下水位的变化受到诸如人工抽水、植被变化、降雨等多种外界因素的干扰,对地下水位的动态变化进行准确预测是一个挑战较大的问题。集对分析^[10]是中国学者冯克勤提出的一个方法,用联系度来描述一个系统中的确定性和不确定性部分,并由此实现系统的评价和决策。万星等^[11]总结了集对分析在水文水资源中的应用前景,其潜在应用主要包括水文预报误差评定、流域年水量平衡状

收稿日期:2012-11-02

基金项目:国家自然科学基金面上资助项目(91125006)

作者简介:何思为(1988—),男,硕士生。研究方向:黑河流域陆面过程模型。E-mail:hesiweide@163.com

* 通信联系人 E-mail: nztong@lzb.ac.cn

况表示、以及相似流域的客观选择。欧源等^[12]在文献[11]的研究基础上,提出了基于秩次进行相似性评价,即秩次集对分析方法,并运用方法预测年径流,表明该方法预测精度较高。王延亭等^[13]对秩次集对分析方法进行了改进,提出了基于加权的秩次集对分析方法,在黑河莺落峡站的预测结果表明该方法精度较高。何菡丹等^[14]把小波方法和秩次集对分析方法相结合,并将其预测能力与单纯的秩次集对分析方法进行了比较,指出和小波方法结合后,其预测能力有明显的提高。但是这些秩次集对分析方法在预报中的应用研究都只考虑一个后续值,也就是每次只预测一个值,这种短序列的预测在实际应用中局限性很大,往往不能满足实际需求。已有研究表明秩次集对分析方法在预测未来一个值时候精度较高,该方法预测较长的一个序列时其精度怎样,误差如何变化,不同的历史集合容量及后续值会对该方法预报精度产生什么影响,这些问题仍然不清楚。为解决上述问题,作者用秩次集对分析方法来分别预测西北干旱地区2个测井的单后续时刻和多后续时刻的地下水位动态变化,通过比较历史集合容量 T 和后续值个数 n 对该方法预测效果的影响及相应的误差变化,深入分析该方法在预测地下水位动态变化时的一些特征,为今后该方法的使用提供一些参考。

1 秩次集对分析方法

秩次集对分析方法^[12]是在集对分析原理^[10]的基础上提出的一种预测手段。秩次集对预测方法应用相似预测的原理,认为未来事物的状态与历史有一定的相似性,用历史的状态来预测未来的状态,其基本原理如下:

设已知时间序列 $X_1, X_2, X_3, \dots, X_n$ 且 X_i 与前 T 个历史值 $X_{i-1}, X_{i-2}, \dots, X_{i-T}$ 相关,将该时间序列滑动生成容量为 T 的集合,分别记为 $A_1, A_2, A_3, \dots, A_{n-T}$,这里的集合 A_i 称为历史集合,每个历史集合对应一个后续值 $X_{T+1}, X_{T+2}, X_{T+3}, \dots, X_n$ 。称集合 $B = \{X_{n-T+1}, X_{n-T+2}, \dots, X_n\}$ 为当前集合,其后续值 X_{n+1} 就是需要预测的值。对 X_{n+1} 采用下述方法预测:在历史集合 $A_1, A_2, A_3, \dots, A_{n-T}$ 中寻找与当前集合 B 相似的集合 A_i 的后续值 X_{T+i} 作为 X_{n+1} 预测值。在历史集合中找到与当前集合相似的一个或者几个集合采用秩次方法。称一个集合中元素的相对大小位次为秩次,首先对历史集合和当前集合做秩变换得到秩次集合 $A_1', A_2', A_3', \dots, A_{n-T}', B'$,再将 $A_1', A_2',$

A_3', \dots, A_{n-T}' 分别与 B' 构成集对 (A_i', B') ,称之为秩次集对。然后根据式(1)计算出各集对的联系度:

$$\mu = \frac{S}{T} + \frac{F}{T}i + \frac{P}{T}j \quad (1)$$

其中: μ 为联系度; T 为历史集合容量; S 为同一性的个数; F 为差异性的个数; P 为对立性的个数,且 $T = S + F + P$; i 为差异不确定系数; j 为对立系数。

最后通过联系度找出 B' 的相似集合 A_i' ,也就是 B 的相似集合 A_i 。

2 数据和方法

2.1 观测资料介绍

选择甘肃省白银市景泰县2个测井地下水位进行模拟研究,其观测每月均有6次,分别是1、6、11、16、21、26日。测井1(1#)是红柳泉观测井(编号40300015),观测资料为1981—2003年,其中缺失1992年数据,在该序列中,由于1984—1985年人工干扰导致数据异常,从观测资料中去掉这2年的数据。测井2(2#)是南滩八队观测井(编号40300036),观测资料为1981—2002年,其中缺失1992年数据。

2.2 研究方法

秩次集对分析方法是一种基于统计的方法,在使用该方法时历史集合容量 T 值的选取一般不同,即 T 的取值和物理现象有关。同时, T 的取值还受到观测资料长度的限制,如果资料长度一定的情况下 T 值越大,相似集合的个数就越少,从而使得预测的精度下降。欧源等对流域年径流进行预测时,推荐 T 取4~6^[12]。为了研究秩次集对分析方法在甘肃省白银市景泰县地下水位预测中 T 取值的变化对预测精度的影响进行了研究,分别取 $T = 3, 4, 5, 6, 8, 10, 13, 15$,对地下水位进行预测,然后比较不同 T 值下的预测误差,采用均方误(RMSE)来表示模拟误差,其计算公式如下:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_{obs,i} - x_{sim,i})^2}{n}} \quad (2)$$

式中, $x_{obs,i}$ 表示观测值, $x_{sim,i}$ 表示预观测值, n 表示序列长度。已有文献[11-13]在应用秩次集对方法时,均只预测未来一个时刻,这种情况称之为单后续值的情形。同时考虑多后续值的情况,即在预测时历史集合和当前集合的后续值个数为 $n(n > 1)$ 个,每完成一次当前集合的相似集合查找之后可以预测当前集合的后续 n 个值。多后续值相对单后续值的优

点是在预测相同长度时间序列的情况下,可以减少预测次数,也就是减少误差累积次数。这种方法在预测相同长度序列的地下水位动态变化时,趋势误差累积次数只有原来的 $1/n$, 减缓了预测中趋势误差的累积。

3 结果与讨论

3.1 单后续值预测

3.1.1 历史集合容量的影响

分别取 T 为 3、4、5、6、8、10、12、15, 分析历史容量 T 和模拟误差之间的关系。对测井 1, 模拟 1981—2002 年每次观测对应的地下水位值, 对测井 2, 模拟 1981-01-01—2001-12-26 年地下水位动态变化, 模拟结果分别如图 1(a)、(b)。通过对 2 个测井地下水位的模拟, 得到历史集合容量 T 与预测误差 $RMSE$ 的关系, 如图 2 所示。

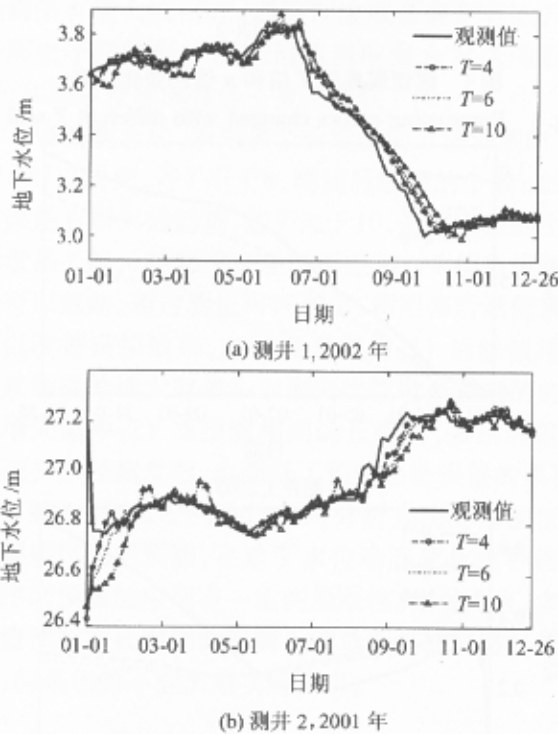


图 1 不同 T 值对应的地下水位模拟值

Fig.1 Simulated groundwater tables with different T values

从图 2 可以看到, 秩次集对分析方法在模拟该地区 2 个测井地下水位时的误差变化趋势一致, 均随着历史集合容量的增多均方误差增大。在历史集合容量小于 8 时候, 均方误差的增大幅度很小, 当样本容量大于 10 时, 由于集合内序列数据太长, 利用秩次方法很难找到相似集合, 即在观测地下水位序列中很难找到变化趋势相似的两段, 因此很多值无法预测, 这导致均方误差急剧增大。在预测地下水位动态变化时, 历史

集合容量 T 值的范围与文献[12] 推荐的使用该方法预测流域年径流时 T 的取值基本一致。

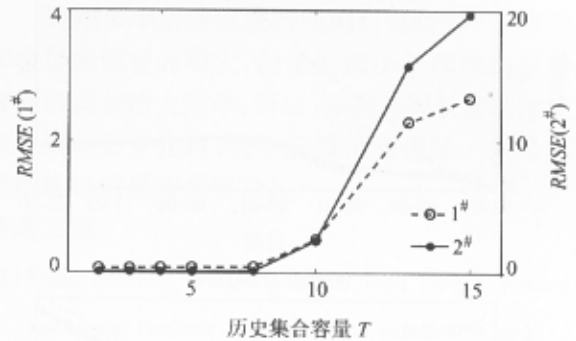


图 2 模拟误差 $RMSE$ 和历史集合容量 T 的关系

Fig.2 Relation between simulation error ($RMSE$) and the historical set size (T)

3.1.2 长序列预测检验

上面情况说明当 T 不大于 10 时, 秩次集对方法在预测地下水位动态变化未来 1 个值时, 精度较高, 但是, 这在实际中的应用价值并不高, 只有能够较精确的预测地下水位在 1 个较长的序列动态变化时 (年际变化), 才对实际生产有指导意义。下面检验秩次集对分析方法预测 1 a 内地下水位变化的能力。

对测井 1, 分别取 $T = 4, 6, 8, 10$, 将 1981-01-01—2002-12-26 的地下水位观测数据作为已知, 模拟 2003 年地下水位动态变化, 对测井 2, 按照和测井 1 一样的方式来模拟 2002 年地下水位的变动, 模拟结果误差变化分别见图 3(a) 和 (b)。从图 3 可以看到, 当用秩次集对方法预测较长的时间序列的时候, 误差累积相当严重, 这主要是由于该方法原理上的缺陷。方法在预测未来值时, 在历史序列中用秩次方法确定与当前集合变化趋势相似的序列, 由于随着预测序列的延长, 当前集合中元素里预测值逐步增多, 观测值减少, 这样, 如果预测值有减小的趋势, 秩次方法就会确定有下降趋势历史序列的后续值来预测当前集合对应的后续值, 如果预测值有增大的趋势, 则秩次方法就会确定有上升趋势历史序列的后续值来预测当前集合对应的后续值 (后面称这种误差为趋势误差)。所以, 用秩次集对分析方法无法预测地下水位未来较长时间的变化。

3.2 多后续值预测

3.2.1 历史集合容量及后续值个数影响

从 3.1 节可以看到, 秩次集对分析方法在预测时候趋势误差累积相当严重, 对长时间序列模拟效果不好。在本小节对多后续值情形进行研究。

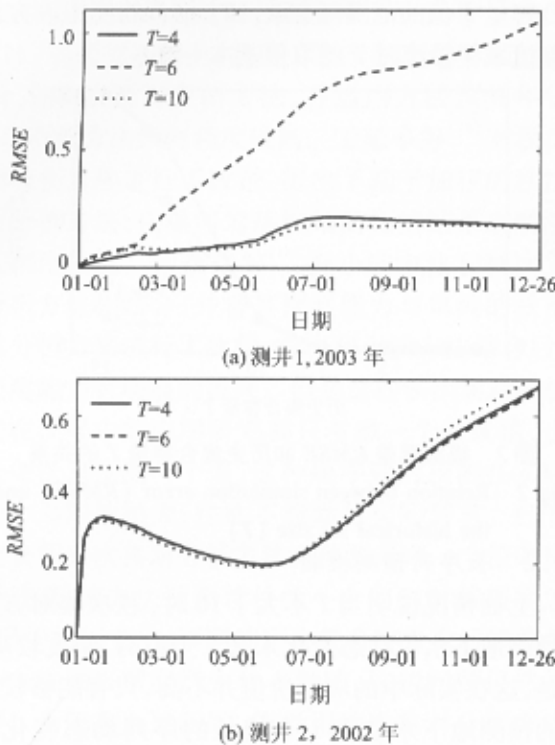


图3 预测误差随预测长度的变化

Fig. 3 Forecasting errors changed with prediction lengths
对测井1, 分别在 $T = 6, n = 2, 4, 6; T = 8, n = 2, 4, 6, 8; T = 10, n = 2, 4, 6, 8, 10$, 总共在 12 种不同的情况下模拟 1981-01-01—2002-12-26 年每次观测对应的地下水水位值, 来分析历史值容量 T 和后续值个数 n 对模拟效果的影响和模拟误差之间的关系, 模拟结果的误差变化见图 4(a)。对测井 2, 采用和测井 1 相同的 T 值变化和 n 值变化及模拟方法, 模拟 1981-01-01—2001-12-26 年地下水水位动态变化, 模拟结果的误差变化见图 4(b)。从图 4 可以看到, 当 T 小于 8 时, 随着后续值的增多, 预测误差逐渐增大, 当 T 为 10 时候, 预测误差主要的决定因素是历史集合容量 T , 与后续值的个数 n 关系很小。

3.2.2 长序列预测验证

比较图 3 和 4, 可以看到当历史集合容量 T 在 8 以内时, 因为后续值的增多而带来的误差累积速率远小于预测次数增多带来的误差累积速率, 于是, 作者尝试用多个后续值模拟地下水水位动态变化的一个长序列。对测井 1 取 $T = 6, n = 2, 3, 4, 5, 6$, 模拟 2003 年的地下水水位变动, 将模拟值与观测值比较, 模拟结果误差变化如图 5(a) 所示。对测井 2 用与测井 1 相同的方法, 预测 2002 年地下水水位动态变化, 模拟结果误差变化如图 5(b) 所示。从图 5 可以看到, 相对于 $n = 1$ 时的误差增长速率, 当 $n > 1$ 时的误差累积速率变小, 模拟效果明显改善。

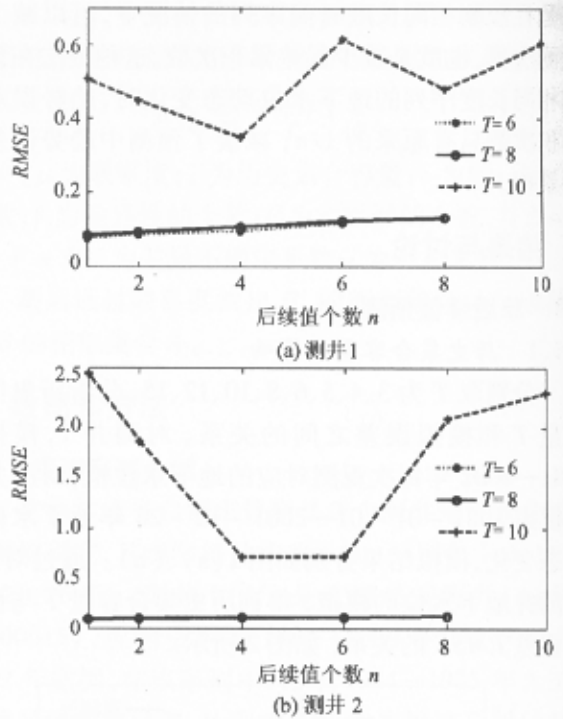


图4 模拟误差随 T 值和 n 值的变化关系

Fig. 4 Forecasting errors changed with different T and n

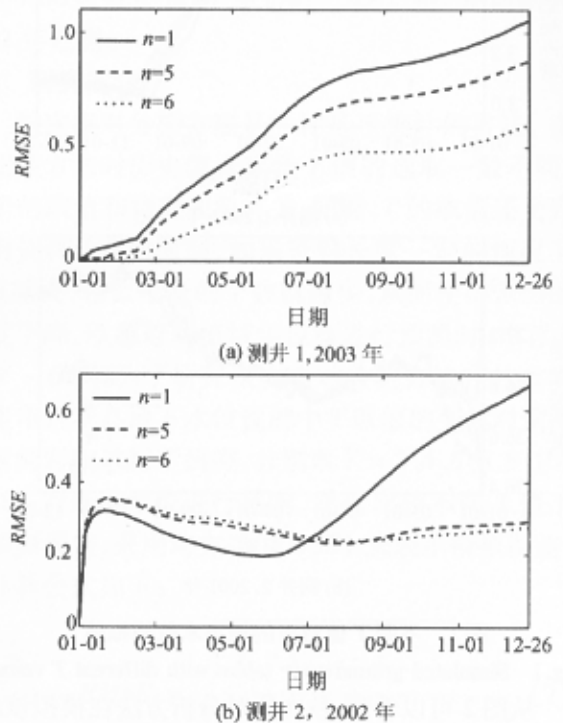


图5 T=6 时, 预测误差与 n 的关系图

Fig. 5 Relation between forecast error and n when T=6

3.3 讨论

从前面的模拟结果和分析可以得出, 秩次集分析方法在预测地下水水位动态变化时单后续值情况下一步预测精度较高, 如图 2, 当 T 小于 8 时, 其误差

基本不随历史集合的容量 T 的变化而变化,但当 T 大于 8 时,其误差随着历史集合的容量 T 的变大急剧增大,这主要是因为集合里元素序列太长,在方法设定时,指定只有当联系度 $\mu \geq 0.5$ 时才认为历史集合与当前集合相似,致使许多当前集合无法在历史集合中找到相似集合,进而无法确定当前集合的后续值。但是在验证过程发现,不论 T 取何值,该方法的误差累积相当严重,所得到的预测趋势总是一直下降或者上升,如图 3 所示,使得所得的模拟值基本不具有对地下水位动态变化的描述能力。这主要是因为方法在预测未来值时,该方法的核心是在历史序列中确定与当前集合中元素序列变化趋势相似的序列,而随着预测序列的延长,当前集合中元素里预测值逐步增多,真实观测值减少,这样,如果预测值有减小的趋势,该方法就会确定有下降趋势历史序列的后续值来预测当前集合对应的后续值,如果预测值有增大的趋势,则该方法就会确定有上升趋势历史序列的后续值来预测当前集合对应的后续值,从而致使预测趋势一直上升或下降。

当改变方法,采用多个后续值的方法预测时,从图 4 可以看到,若 T 小于 8,则随后续值的个数 n 的增多误差有增大的趋势,若 T 大于 10,造成误差的主要因素是 T 值, n 值增多产生的误差相对 T 值造成的误差可以忽略。通过验证可以看出,利用多后续值预测可以改善模拟效果,这主要是因为:1) 后续值增多带来的误差增大速率小于预测次数增多带来的误差增大速率;2) 当预测相同的长度时,多后续值预测减少了预测次数,也降低了预测趋势误差的累积。

综上,可以看到秩次集对分析方法的精度仅限于很短序列的预测,在地下水位动态变化这种需要长序列预测的应用有一定的局限性。相对而言,多后续值预测在长序列的预测上较单后续值有略好的精度,但是仍然不能满足实用要求。

4 结 论

根据研究,秩次集对分析方法在预测地下水位变化上有很大的局限性,在预测上主要表现出以下一些特点:

1) 秩次集对方法在预测该地区的地下水位变化时,当历史集合容量 T 小于 8 时,预测误差基本不随着 T 变化,当历史集合的容量 T 大于 10 时,预测误差随 T 值的增大急剧增大。

2) 秩次集对分析方法在预测未来 1 个值时,有较好的预测精度,但是,如果在预测长时间序列时,

预测误差累积相当严重,该方法预测得到的值不再具有对地下水位动态变化的描述能力。

3) 用多个后续值进行预测时,虽然后续值的增多会带来误差的增大,但是远远小于预测次数增多带来的误差增大速率,所以,预测相同长度的地下水位序列动态变化时,较单后续值有优势。但是精度仍然难以满足实用要求。

参考文献:

- [1] Wang Jinsheng, Wang Changshen, Teng Yanguo. Review on assessment methods of groundwater sustainable yield [J]. Journal of Hydraulic Engineering, 2006(5): 525 - 533. [王金生,王长申,滕彦国.地下水可持续开采量评价方法综述[J].水利学报,2006(5):525 - 533.]
- [2] Zheng Dan, Li Weihong, Chang Yapeng, et al. Relations between groundwater and natural vegetation in the arid zone [J]. Resources Science, 2005(4): 160 - 167. [郑丹,李卫红,陈亚鹏,等.干旱区地下水与天然植被关系研究综述[J].资源科学,2005(4):160 - 167.]
- [3] Yang Zeyuan. Study on supergene ecological effect excited by groundwater and its evaluation [D]. Xi'an: Chang'an University, 2004. [杨泽元.地下水引起的表生生态效应及其评价研究[D].西安:长安大学,2004.]
- [4] Tao Yuezan, Jiang Ling. Influence of groundwater discharge on decay process of contamination concentration in river flow [J]. Journal of Hydraulic Engineering, 2008(2): 245 - 249. [陶月赞,蒋玲.地下水补给对河流污染物浓度衰减过程的影响[J].水利学报,2008(2):245 - 249.]
- [5] Ding Zhenyu. Groundwater recharge and evolution in shiyang river basin and tengger desert, Northwestern China [D]. Lanzhou: Lanzhou University, 2010. [丁贞玉.石羊河流域及腾格里沙漠地下水补给过程及演化规律[D].兰州:兰州大学,2010.]
- [6] Wu Yanqing. Multivariable autoregression model for groundwater level forecasting of multiple wells [J]. Geotechnical Investigation and Surveying, 1993(2): 23 - 27. [仵彦卿.多井地下水位预测的多变量回归模型[J].工程勘察,1993(2):23 - 27.]
- [7] Chen Hanjun, Yang Xue, Huang Dongwei. Application of time series to groundwater level prediction [J]. Journal of

- Tianjin University of Technology, 2008(2): 8 - 10. [陈汉军, 杨雪, 黄东卫. 时间序列在地下水位预测中的应用[J]. 天津理工大学学报, 2008(2): 8 - 10.]
- [8] Jiang Bo, Wang Feng, Luo Liyan, et al. Application of artificial neural network model in forecasting groundwater level of dry season in Baicheng area[J]. Water Resources & Hydropower of Northeast China, 2010(12): 19 - 20. [姜波, 王锋, 罗丽燕, 等. 人工神经网络模型在白城地区枯季地下水位预测中的应用[J]. 东北水利水电, 2010(12): 19 - 20.]
- [9] Wu Dongjie, Wang Jinsheng, Teng Yanguo. Application of wavelet decomposition and wavelet transform method to forecasting of groundwater regime[J]. Journal of Hydraulic Engineering, 2004(5): 39 - 45. [吴东杰, 王金生, 滕彦国. 小波分解与变换法预测地下水位动态[J]. 水利学报, 2004(5): 39 - 45.]
- [10] 冯克勤. 集对分析及其初步应用[M]. 杭州: 浙江科学技术出版社, 2000: 1 - 198.
- [11] Wan Xing, Wang Wensheng, Ding Jing. Set pair analysis and its application to hydrology and water resources[J]. Advances in Science and Technology of Water Resources, 2006(4): 9 - 11. [万星, 王文圣, 丁晶. 集对分析在水文水资源中的应用[J]. 水利水电科技进展, 2006(4): 9 - 11.]
- [12] Ou Yuan, Zhang Qiong, Wang Wensheng, et al. An annual flow forecast model based on rank set pair analysis[J]. Yangtze River, 2009(3): 63 - 65. [欧源, 张琼, 王文圣, 等. 基于秩次集对分析的年径流预测模型[J]. 人民长江, 2009(3): 63 - 65.]
- [13] Wang Yanting, Wang Jianqun, Zhang Yujie. Annual runoff forecasting model based on weighted rank set pair analysis method[J]. Water Resources and Power, 2012(3): 17 - 19. [王延亭, 王建群, 张玉杰. 基于加权秩次集对分析法的年径流预报模型[J]. 水电能源科学, 2012(3): 17 - 19.]
- [14] He Handan. Hydrologic time series prediction model based on wavelet de-noising and rank set-pair analysis[D]. Nanjing: Nanjing University, 2012. [何菡丹. 基于小波消噪与秩次集对分析的水文时间序列预报模型[D]. 南京: 南京大学, 2012.]

(编辑 张琼)