

# 基于随机森林的黑河中游地下水埋深变化及成因

杨光, 粟晓玲

(西北农林科技大学 水利与建筑工程学院, 陕西 杨凌 712100)

**摘要:**研究黑河中游地下水埋深变化规律及其成因,可为该流域地下水合理开发利用提供参考。根据黑河中游33个观测井1985—2010年的地下水埋深资料,利用地统计学方法分析了地下水埋深变化规律;采用随机森林法研究了埋深变化的成因。结果表明:黑河中游地下水埋深自南向北减少;甘州区和临泽县的地下水埋深变化趋势分别在2001年和2004年前呈增大趋势,之后又逐年减小,高台县北部地区埋深呈持续减小趋势,而东部地区呈增大趋势;地统计学分析表明地下水埋深具有中等的空间相关性,且空间各向异性有增强的趋势;影响黑河中游地下水埋深变化的主要因素是渠系水利用系数、地下水开采量、地表水引用量及蒸发能力。

**关键词:**黑河中游;地下水埋深;随机森林法;地统计学

**中图分类号:**P641.8;S273.4

**文献标识码:**A

**文章编号:**1005-3409(2017)01-0109-06

## Change of Groundwater Depth and Its Causes in Middle Stream of the Heihe River Basin Based on the Random Forest

YANG Guang, SU Xiaoling

(College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China)

**Abstract:** In order to provide useful references for the reasonable development and utilization, the distribution characteristics of groundwater depth and its causes in middle stream of the Heihe River Basin was examined. Based on the local groundwater depth data of 33 observation wells from 1985 to 2010, the spatiotemporal distribution characteristics of the groundwater depths were analyzed by using the geostatistics. The random forest method was used to investigate the change causes of groundwater depth. The results showed that the groundwater depths of the middle stream of the Heihe River Basin reduced from the south to the north. Before 2001 in Ganzhou District and 2004 in Linze County, the groundwater depths had an increase trend, respectively, afterwards both reduced. The groundwater depth of the northern Gaotai County presented the continuous decrease trend, but groundwater depth of the eastern Gaotai region increased. Moreover, according to the analysis of geostatistics, the sample of the groundwater depths had spatial correlations with middle-level, and its spatial anisotropy had enhancement tendency. In addition, the main factors influencing the groundwater depth in the middle stream of the Heihe River Basin were the utilization coefficient of canal system water, the exploitation quantity of the groundwater, surface water diversion quantity and evaporative capacity.

**Keywords:** middle stream of the Heihe River Basin; groundwater depth; the random forest method; geostatistics

地下水系统的时空分布与演变规律,既受天然条件,如地质环境、气候环境、水环境和生态环境等因子控制,又受农业灌溉、工业化、城市化等人类各种经济活动的影响<sup>[1]</sup>,地下水的补给和排泄发生动态变化,从而引起地下水埋深的时空演变。我国北方地区,降雨稀少,水资源短缺,地下水过度开采,已导致生态环境恶化,如石羊河流域、黑河流域。研究地下水埋深的时空变化规律和变化成因,对合理利用地下水以及

保障区域社会经济持续发展具有重要的指导意义。

目前对于地下水位空间变异规律的研究多采用地统计学方法,如周剑等<sup>[2]</sup>采用地统计学中的普通克里金方法研究黑河中游土地利用变化对地下水位时空变异性的影响。魏智等<sup>[3]</sup>将地统计理论应用于黑河中游分水前后的地下水位及储量的比较分析;对地下水位(或埋深)变化的成因分析方法主要有主成分分析法,如张素芳<sup>[4]</sup>采用主成分分析法研究影响咸

收稿日期:2015-12-22

修回日期:2016-03-10

资助项目:国家自然科学基金(91425302;51279166)

第一作者:杨光(1990—),女,辽宁抚顺人,硕士研究生,研究方向为地下水演变与可持续利用。E-mail:1007153730@qq.com

通信作者:栗晓玲(1968—),女,四川开江县人,教授,博士,博士生导师,主要从事水资源规划与管理研究。E-mail:suxiaoling17@126.com

阳地下水动态变化的驱动因子。陈伏龙等<sup>[5]</sup>将主成分分析法运用于分析影响莫索湾灌区地下水埋深的因素。但主成分分析法存在不能单一地去评价某一研究对象的某种性能的问题<sup>[6]</sup>。甚至有学者指出主成分分析法在综合评价函数中对各主成分使用贡献率加权是错误的<sup>[7]</sup>。本文采用的随机森林法(random forest)具有在运算量没有显著增加的前提下提高预测精度的优点,被誉为目前最好的算法之一<sup>[8]</sup>。近年来,被广泛应用于医学临床及生物研究<sup>[9-10]</sup>。但应用于地下水方面的研究较少。Peters 等<sup>[11]</sup>用随机森林法分析地下水位在不同种类植被下的响应;Baudron 等<sup>[12]</sup>运用随机森林法对提取的地下水按离子成分进行分类,精度达到 90% 以上。

黑河中游是水资源的耗散区,其对地下水的利用方式决定了黑河流域中游地下水的补排方式和生态环境的演变<sup>[13]</sup>。本文基于 ArcGIS 地统计分析方法以及随机森林法分析黑河中游地下水埋深的时空演变规律及其变化成因。

## 1 研究区概况及研究方法

### 1.1 研究区概况

黑河流域是我国西北干旱区较大的内陆河流域之一,位于祁连山和河西走廊的中段。研究区黑河中游流域位于 99°20′—100°50′E, 38°40′—39°59′N, 其行政区分为甘州区、临泽县、高台县,总面积 1.07 万 km<sup>2</sup>。黑河流域中游盆地开阔平缓,地势南高北低,东部略高于西部,在南部祁连山麓盆地边缘海拔为 1 800~2 000 m,北部黑河沿岸海拔为 1 400~1 500 m,面积约为 6 375.86 km<sup>2</sup><sup>[14]</sup>。研究区内多年平均降水量 58~216 mm,多年平均潜在蒸发能力为 860~1 160 mm,地下水类型主要有潜水及承压水,潜水埋藏浅且易于开采,是农业用水的主要来源。承压水一般埋深约在 100 m 以下,不易开采。本次研究对象主要为潜水。

2000 年以前,黑河中游人口和灌溉面积大幅度增加。为满足用水需求,中上游大量截引或滞留河水。因此进入下游水量常年不足,导致河道断流加剧,湖泊干涸<sup>[15]</sup>。为改善黑河下游干旱状况,国务院审批通过了黑河中下游分水方案[水政资(1997)496 号],2001 年 2 月确定实施黑河近期治理,通过灌区节水改造、退耕还林还草和经济结构调整等措施,实现当黑河上游来水 15.8 亿 m<sup>3</sup> 时,向下游下泄 9.5 亿 m<sup>3</sup> 水的流域分水目标。随着分水方案的实施,中游地下水开采量逐年增加,地下水的补排条件也发生了变化。

### 1.2 研究方法

1.2.1 数据来源及处理 地下水埋深数据来源于黑

河中游 1985—2010 年 33 眼地下水位动态观测井,其空间分布见图 1。

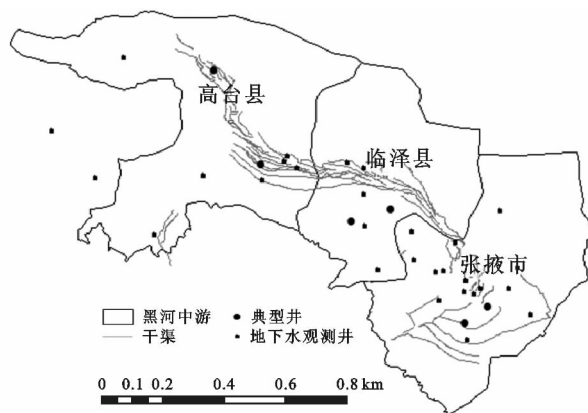


图1 研究区地下水观测井空间分布

1.2.2 地统计学方法 克里金插值是地统计学的主要内容之一,是以变异函数理论和结构分析为基础,在有限区域内对区域化变量进行无偏最优估计的一种方法<sup>[16]</sup>。

1.2.3 随机森林法 随机森林(random forest)是由 Breiman 和 Cutler 在 2001 年借鉴贝尔实验室的 Ho 所提出的把分类树组合成随机森林的方法<sup>[17]</sup>。本文以地下水埋深为因变量,将 6 个影响地下水埋深演变因素作为自变量,计算影响因素的重要性,从而获得黑河中游地下水埋深演变的主要因素。

随机森林计算步骤<sup>[18]</sup>: (1) 根据 Bootstrap 自助法最原始的样本里的数据中有放回的随机抽出  $N$  (如 70%) 个研究的样本后,作为  $N$  棵树,树的数量即为  $N$ 。 (2) 在不同树生成的时候,在研究的全部变量中平均选取变量数量为  $M$ ,随后在  $M$  中挑选分类能力最强的一个变量,分类原理见图 2。 (3) 在上述抽取样本的过程中没有被抽中的剩余数据 (如 30%) 作为本研究的测试集,以此来对每棵树的相应性做出一定的评价。 (4) 由自变量的重要性评分来分析地下水变化的成因。 (5) 查看模型的标准残差值及预测误差值。随机森林法可以通过 R 语言等工具实现。R 语言是一个自由、免费、开源的软件,是一个具有强大统计分析和优秀统计制图功能的统计软件<sup>[19]</sup>。R 语言中有 Random Forest 软件包可以运行随机森林。

## 2 结果与分析

### 2.1 地下水埋深年际变化规律

根据至黑河干渠垂直距离近、较近、远以及数据的完整程度,选取高台县罗成乡及正远乡、甘州区小满乡及梁家墩乡、临泽县东南角及新华乡的观测井作为典型井。

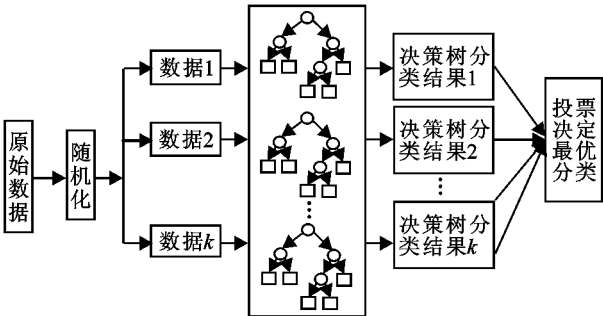


图 2 随机森林分类原理

从图 3 分析可知,位于高台县北部的罗成乡呈减小趋势,而东部的正远乡地下水埋深的变化趋势则大幅度增加。甘州区 1985—2001 年地下水埋深逐渐增

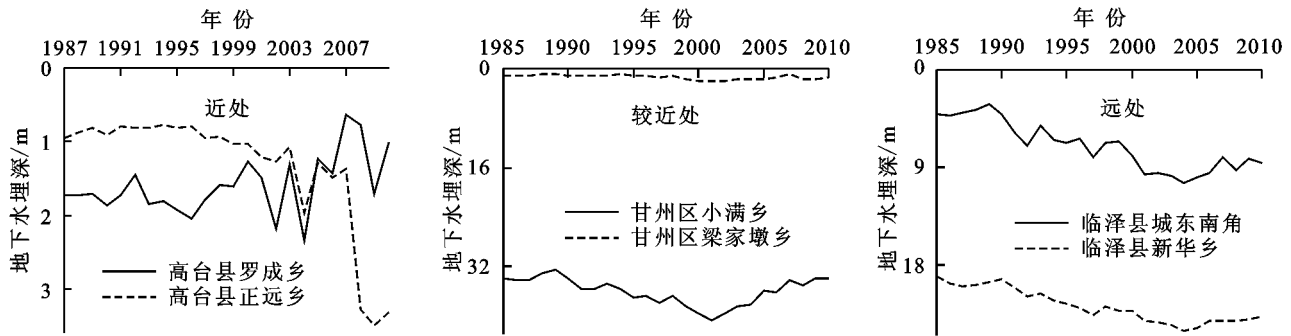


图 3 典型井地下水埋深年际变化

表 1 1985—2010 年地下水埋深统计分析成果

年份	样本数/眼	最小值/m	最大值/m	平均值/m	标准差	偏度系数	峰度系数
1987	26	0.79	274.39	36.145	70.050	-0.481	1.936
1992	22	0.15	249.75	42.451	74.527	0.317	1.808
1998	31	0.29	251.38	34.298	64.459	0.332	1.925
2001	33	0.57	251.93	32.981	62.521	0.322	1.979
2004	29	0.31	238.26	28.532	52.550	0.097	2.186
2010	27	0.01	239.23	32.343	59.888	0.294	2.306

注:偏度系数及峰度系数是对原数据取对数处理后的值。

2.2.2 地下水埋深空间变异特性 以 2004 年为例分别采用球状模型、指数模型及高斯模型模拟灌区地下水埋深,各模型交叉验证结果见表 2。由表 2 可知,球状模型拟合得较好。球状模型不同半方差函数模型内插参数设置见表 3。地下水位的空间相关性可以根据块金值与基台值比值的大小进行划分,块金系数小于 25% 时,空间相关性强;在 25%~75% 时,空间相关性中等;大于 75% 时,空间相关性弱<sup>[21]</sup>。由表 3 可知,除 1987 年地下水埋深为较强的空间相关性,其他典型年的块金系数为 25%~75%,说明具有中等的空间相关性;长轴所在的角度大致为 100°,短轴变程与长轴变程比值逐年减小,说明空间各向异性略有增强。

2.2.3 地下水埋深的空间分布 结合典型井地下水埋深年际变化,参考数据的完整程度,将典型年设为

加,而在 2001—2010 年埋深逐年减少。临泽地区在 1985—2004 年间地下水埋深持续增加,在 2004 年埋深达到最大值,此后地下水埋深略有减小。

2.2 地下水埋深的空间分布特征及变异特性

2.2.1 样本统计分析 以 2004 年为例,在经过对数变换后的正态 QQ plot 中数据靠近表示正态分布的直线且离群现象不太明显,基本满足正态分布。对黑河中游地下水埋深统计分析,结果见表 1。由于数据不完整,导致典型年的井数不一致。经对数变换后,各年份的偏度系数接近于 0,同时峰度系数接近 2,表明样本数据在空间上基本满足对数正态分布<sup>[20]</sup>。

1987 年、1992 年、1998 年、2001 年、2004 年、2010 年。利用 ArcGIS 软件中克里金法对黑河中游地下水埋深进行插值,绘制典型年地下水埋深空间分布图(图 4)。由图 4 可知,黑河中游地下水埋深自南向北减小。高台县北部地下水埋深在波动减小,而东部明显增大。用图 3(典型井地下水埋深的年际变化)校核图 4 的准确性,比较可知变化趋势一致。

表 2 2004 年不同模型交叉验证统计值

克里金模型	误差均值	均方根误差	平均标准误差	标准均方根预测误差
球状模型	-2.366	50.848	14.287	9.758
指数模型	-2.760	54.397	12.828	9.833
高斯模型	-2.742	53.784	12.393	10.173

2.3 黑河中游地下水埋深变化成因

2.3.1 主要影响因素的变化过程 绿洲区地下水系统的补给项主要来源于降水量、河道及渠系渗漏补

给、灌溉入渗补给,排泄项包括地下水开采量及潜水蒸发。本文选择降水量、地表水引用量、地下水开采量、渠系水利用系数、蒸发能力及莺落峡来水量 6 项因素,各因素的变化过程见图 5。甘州区、临泽县、高台县的降水量年际变化波动较大,2000 年后有上升趋势;地表水引用量在逐年下降,甘州区下降明显;地下水开采量呈增加趋势,甘州区和高台县 1998 年以后增加显著,这是由于为应对分水带来的“政策缺水”

和更加严峻的新一轮水资源危机,采取遍地开花式的集中打井灌溉<sup>[22]</sup>,从 2004 年起甘州区开采量又有所下降;渠系水利用系数整体呈增加趋势,其中临泽县和高台县波动较大,原因是渠道老化失修以及渠道冻胀破坏严重<sup>[23-24]</sup>,后来加以修复;3 个县区的蒸发能力(用参考作物蒸发蒸腾量)呈波动上升趋势;莺落峡来水量的年际变化较大,波动频繁,2001 年之后明显增加。

表 3 不同半方差函数模型内插参数设置

年份	块金值	部分基台值	基台值	块金系数/%	长轴变程/m	短轴变程/m	角度	比值
1987	0.111	0.409	0.520	21.3	744.827	507.221	99.23	0.681
1992	0.092	0.254	0.346	26.6	1388.505	788.503	99.48	0.568
1998	0.297	0.399	0.696	42.7	1192.358	399.690	99.59	0.335
2001	0.311	0.290	0.601	51.7	1278.197	428.566	100.23	0.335
2004	0.328	0.451	0.779	42.1	1081.477	361.557	100.36	0.334
2010	0.339	0.465	0.804	42.2	1882.808	633.273	100.45	0.336

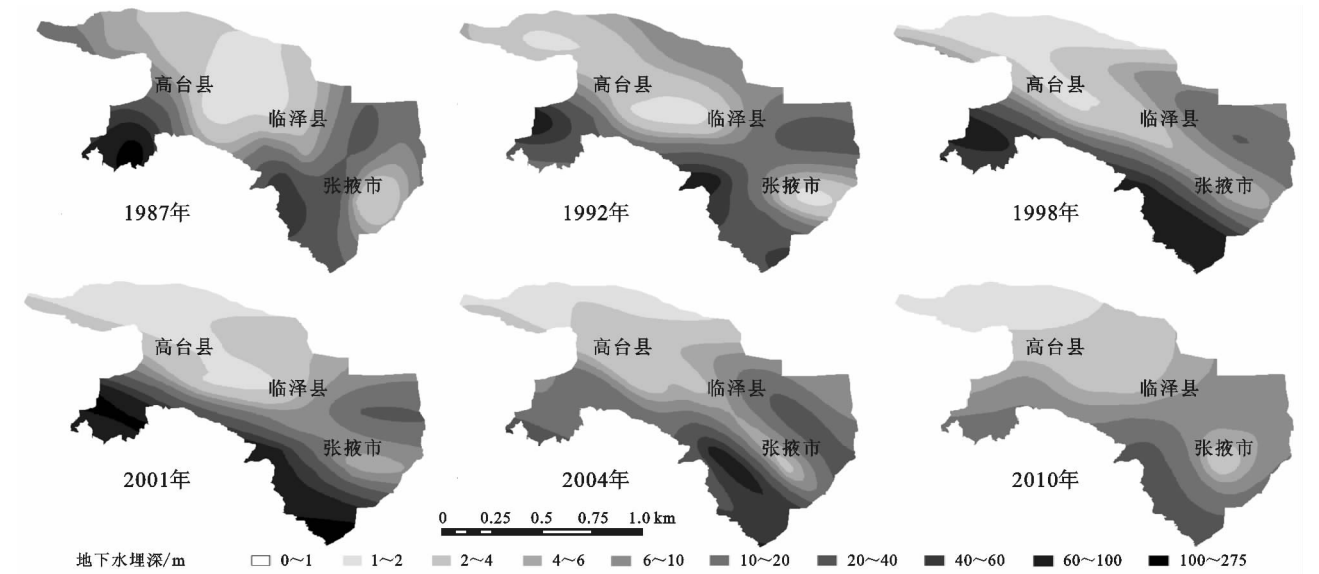


图 4 黑河中游地下水埋深空间分布

2.3.2 重要性评分 利用随机森林模型获得影响因素的重要性排序,结果见表 4。由表 4 可知,各县区地下水埋深变化的影响因素不完全相同,甘州区地下水埋深呈增加趋势的主要原因是蒸发能力的增加、渠系水利用系数的提高以及地表水引水量的减少;临泽县地下水埋深变化的原因是渠系水利用系数的提高、地下水开采量的增加以及地表水引水量的减少;高台县地下水埋深变化的原因是地下水开采量的增加、蒸发能力的增加以及地表水引用量的减少。但总体上渠系水利用系数、地下水开采量、地表水引用量及蒸发能力是影响黑河中游地下水埋深演变的主要因素。

2.3.3 随机森林法预测值与实测值的比较 基于随机森林法对黑河中游地下水埋深进行预测,预测值与实测值拟合效果较好(图 6),预测误差均在 7%以内,标准残差值较小,甘州区、临泽县和高台县分别为

0.633,1.234,0.034 m。说明随机森林法的预测能力较强,结果可信。

表 4 黑河中游地下水演变驱动因素的重要性评分

因素	甘州区		临泽县		高台县	
	重要性评分	排名	重要性评分	排名	重要性评分	排名
降水量	1.557	6	2.340	5	0.087	6
地表水引用量	2.730	3	3.959	3	0.154	3
莺落峡来水量	1.598	5	2.828	4	0.096	5
蒸发能力	3.255	1	1.604	6	0.279	2
渠系水利用系数	2.901	2	12.234	1	0.151	4
地下水开采量	1.993	4	7.776	2	0.363	1

注:甘州区、临泽县及高台县 3 个地区观测井的年平均地下水埋深分别作为自变量,3 个地区的井数分别为 4,4,3 眼井。图 6 中埋深实测值的计算同理。

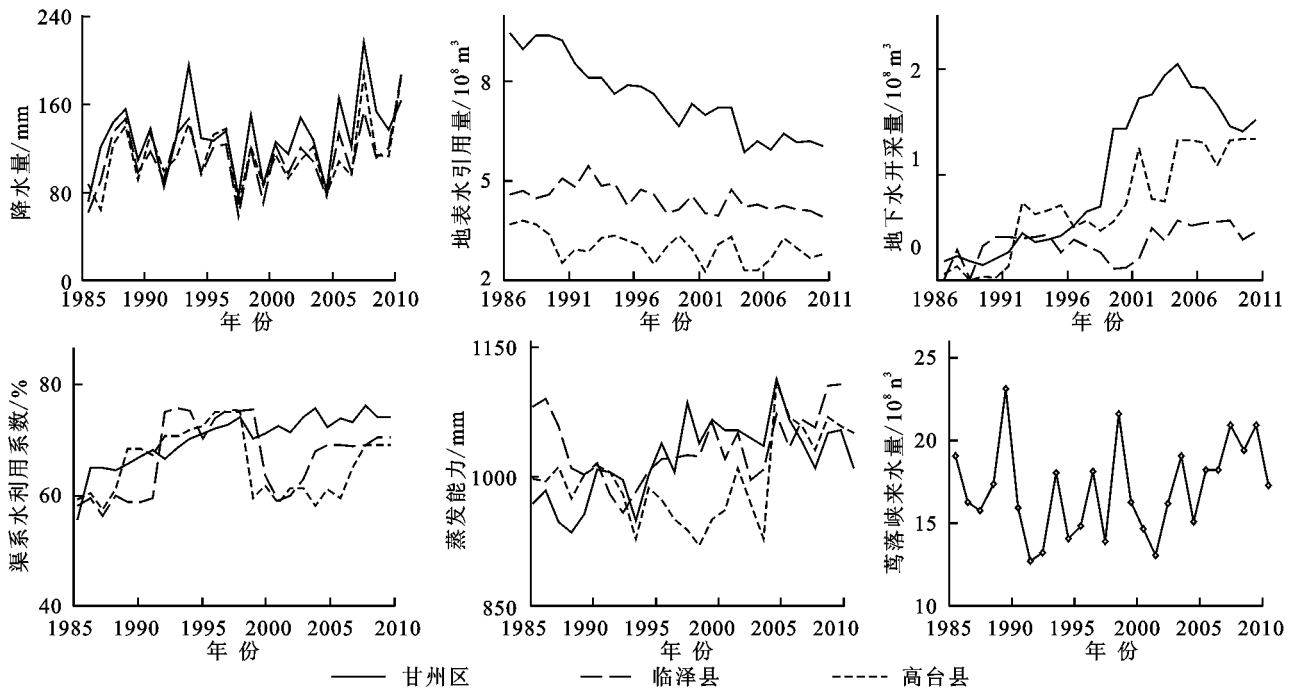


图 5 影响因素年际变化

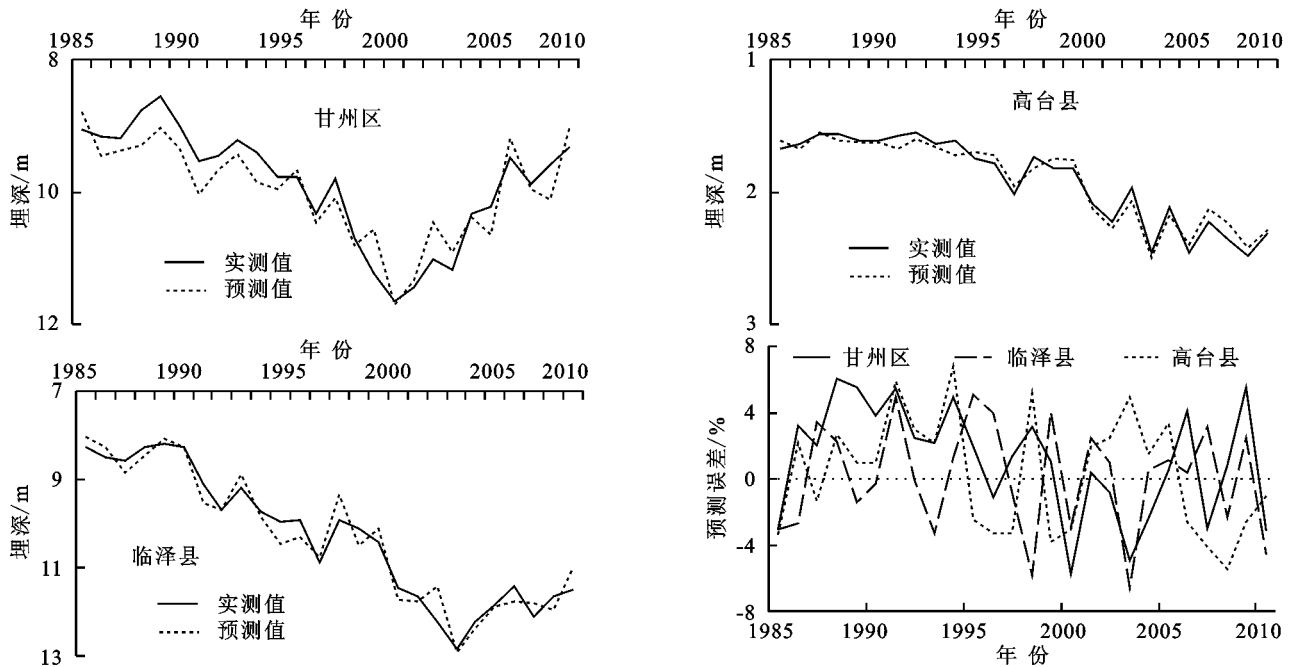


图 6 预测值与实测值的比较及预测误差

### 3 结论

(1) 黑河中游地下水埋深年际变化在各县区不同。高台县部分井的地下水埋深持续减小,另一部分则相反。甘州区在 2001 年埋深达到最大值,临泽县在 2004 年埋深达到最大值,此后地下水埋深又逐年减少。

(2) 球状模型能较好拟合黑河中游各年地下水埋深,空间分布表明埋深自南向北减少,高台县北部地下水埋深波动减小,东部明显增大。地下水埋深样本具有中等的空间相关性,样本空间呈现几何各向异性,且空间各向异性呈增大趋势。

性,且空间各向异性呈增大趋势。

(3) 影响黑河中游地下水埋深变化的主要因素为渠系水利用系数、地下水开采量、地表水引水量及潜在蒸发能力。随机森林法模拟地下水埋深的拟合效果较好,模拟能力较强,结果可信。

#### 参考文献:

- [1] 张光辉,刘少玉,谢悦波,等.西北内陆黑河流域水循环与地下水形成演化模式[M].北京:地质出版社,2005.
- [2] 周剑,李新,王根绪,等.黑河流域中游地下水时空变异性分析及其对土地利用变化的响应[J].自然资源学报,2009,24(3):498-506.

- [3] 魏智,金会军,蓝永超,等.基于Kriging插值的黑河分水后中游地下水资源变化[J].干旱区地理,2009,32(2):196-203.
- [4] 张素芳.基于主成分分析法的咸阳市区地下水动态驱动因子分析[J].水利与建筑工程学报,2012,10(2):126-129.
- [5] 陈伏龙,郑旭荣,何新林,等.莫索湾灌区1998—2007年地下水埋深变化及影响因素[J].武汉大学学报:工学版,2011,44(3):317-320.
- [6] 包磊,雷进生,刘琼.基于主成分分析法的地铁系统安全综合评价[J].三峡大学学报:自然科学版,2011,33(4):57-59.
- [7] 王学民.对主成分分析中综合得分方法的质疑[J].统计与决策,2007(8):31-32.
- [8] Iverson L R, Prasad A M, Matthews S N, et al. Estimating potential habitat for 134 eastern US tree species under six climate scenarios[J]. Forest Ecology & Management, 2008, 254(3):390-406.
- [9] 洪燕珠,周昌乐,张志枫,等.基于随机森林法的慢性疲劳证候要素特征症状的选择[J].中医杂志,2010,51(7):634-638.
- [10] 孙雪莲,舒清杰,欧光龙,等.基于随机森林回归模型的思茅松人工林生物量遥感估测[J].林业资源管理,2015(1):71-76.
- [11] Peters J, De Baets B, Samson R, et al. Modelling ground-water-dependent vegetation patterns using ensemble learning[J]. Hydrology and Earth System Sciences Discussions, 2007,4(5):3687-3717.
- [12] Baudron P, Alonso-Sarria F, García-Aróstegui J L, et al. Identifying the origin of groundwater samples in a multi-layer aquifer system with Random Forest classification[J]. Journal of Hydrology, 2013,499(9):303-315.
- [13] 赵洁,徐宗学,周剑.黑河中游过去20年地下水位空间变异性分析[J].干旱区资源与环境,2011,25(8):172-178.
- [14] 李江,郝新梅,范琳琳,等.黑河中游绿洲地下水位空间变异性研究[J].水力发电学报,2015,34(11):106-115.
- [15] 李启森,赵文智,冯起,等.黑河流域及绿洲水资源可持续利用理念及对策[J].自然资源学报,2005,20(3):370-377.
- [16] 侯景儒,黄竞先.地质统计学的理论与方法[M].北京:地质出版社,1990.
- [17] 李欣海.随机森林模型在分类与回归分析中的应用[J].应用昆虫学报,2013,50(4):1190-1197.
- [18] 凌俐,陈丽,谢兆莉.随机森林法用于分析阿糖胞苷致不良反应发生的影响因素[J].中国药房,2015,26(8):1091-1093.
- [19] 闫朝晖. R 软件在多元统计分析教学中的应用研究[J].科技创新导报,2011(1):157-158.
- [20] 韩业珍,魏晓妹,李立.基于地统计学的地下水位时空变异特征研究[J].人民黄河,2010,32(5):52-53.
- [21] 胡晓利,卢玲.黑河中游张掖绿洲地下水时空变异性分析[J].中国沙漠,2009,29(4):777-784.
- [22] 米丽娜,肖洪浪,朱文婧,等.1985—2013年黑河中游流域地下水位动态变化特征[J].冰川冻土,2015,37(2):461-469.
- [23] 李吉超,缪建新.临泽县发展节水灌溉途径探讨[J].甘肃农业,2002(4):40-41.
- [24] 张汉军.聚苯乙烯保温板在渠道衬砌中防冻胀应用技术[J].甘肃水利水电技术,2003,39(4):294-295.

~~~~~  
(上接第108页)

- [12] 周生贤.国务院关于水污染防治工作进展情况的报告:2008年12月24日在第十一届全国人民代表大会常务委员第六次会议上[J].中华人民共和国全国人民代表大会常务委员公报,2009(1):74-78.
- [13] 杨艳春,闫莉,程伟.黄河流域非点源污染估算研究[J].人民黄河,2011,33(11):66-68.
- [14] 国家环境保护总局,国家质量监督检验检疫总局.地表水环境质量标准 GB3838—2002[S].北京:中国环境科学出版社,2002.
- [15] 李怀恩.估算非点源污染负荷的平均浓度法及其应用[J].环境科学学报,2000,20(4):397-400.
- [16] 刘洁,陈晓宏,周纯,等.非点源污染在东江河流环境中的贡献比例估算[J].中国人口·资源与环境,2014,24(S1):79-82.
- [17] 张富,姚进忠,雷升文,等.甘肃省水土保持综合治理效益研究[M].郑州:黄河水利出版社,2014.
- [18] 应兰兰,侯西勇,路晓,等.我国非点源污染研究中输出系数问题[J].水资源与水工程学报,2010,21(6):90-95,99.
- [19] 廖义善,卓慕宁,李定强,等.基于“径流—地类”参数的非点源氮磷负荷估算方法[J].环境科学学报,2014,34(8):2126-2132.
- [20] 蒋昌波,楚贝,胡世雄,等.涟水流域非点源污染特征[J].长沙理工大学学报:自然科学版,2011,8(4):64-70.
- [21] 张皓天,张弛,周慧成,等.基于SWAT模型的流域非点源污染模拟[J].河南大学学报:自然科学版,2010,38(6):644-650.
- [22] 贺缠生,傅伯杰,陈利顶.非点源污染管理及控制[J].环境科学,1998,19(5):87-91.
- [23] 唐莲,白丹.农业活动非点源污染与水环境恶化[J].环境保护,2003(3):18-20.
- [24] 卜红梅,王岑,张全发.汉江上游金水河流域非点源污染及控制[J].水土保持学报,2009,23(4):33-38.
- [25] 胥彦玲,王苏舰,李怀恩.土地覆被变化对流域非点源污染的影响研究:以黑河流域为例[J].水土保持研究,2010,17(3):250-253.
- [26] 李俊然,陈利顶,郭旭东,等.土地利用结构对非点源污染的影响[J].中国环境科学,2000,20(6):506-510.
- [27] 孙娟,顾霜妹,李强坤.水土流失与农业非点源污染[J].水利科技与经济,2008,14(12):963-965.