

DOI: 10.13476/j.cnki.nsbdqk.2022.0088

康玮, 曹文庚, 徐丽霞, 等. 基于 k -means 聚类与泰森去丛聚的代表性地下水位计算方法[J]. 南水北调与水利科技(中英文), 2022, 20(5): 876-885. KANG W, CAO W G, XU L X, et al. Calculation method of representative groundwater level based on k -means cluster analysis and Thiessen de-clustering[J]. South-to-North Water Transfers and Water Science & Technology, 2022, 20(5): 876-885. (in Chinese)

基于 k -means 聚类与泰森去丛聚的代表性地下水位计算方法

康玮^{1,2}, 曹文庚^{1,3}, 徐丽霞⁴, 南天^{1,3}, 高媛媛⁵, 聂子一²

(1. 中国地质科学院水文地质环境地质研究所, 石家庄 050061; 2. 河北地质大学, 石家庄 050031;

3. 京津冀平原地下水与地面沉降国家野外科学观测研究站, 石家庄 050061;

4. 涞源县水利局, 河北 保定 074099; 5. 水利部南水北调规划设计管理局, 北京 100038)

摘要: 在总结前人常用的区域水位变化值计算方法的基础上, 结合 k -means 聚类分析、泰森多边形去丛聚和水均衡法等方法, 提出一种基于群井多年连续观测水位数据和储变量数据求解代表性地下水位的新方法, 并在数值模拟算例与实例研究区分别进行了计算和验证。结果表明: 在算例中使用本方法得到的代表性地下水位结果与传统水均衡法计算结果相比准确度可达 90.5%; 在实例计算中使用定兴县 8 口井连续 5 年水位数据和储变量数据计算得到定兴县 2019 年代表性地下水位变化值为 0.16 m, 相较水均衡法计算结果新方法的准确度达到 93.3%, 算例与实例结果均较为准确, 计算结果可用于代表性地下水位变化值快速、科学表征, 在实际工作中能够有效简化代表性地下水位变化的计算工作并为结果提供科学依据。

关键词: 地下水动态; k -means 聚类; 泰森多边形; 代表性水位; 超采治理评估

中图分类号: TV213; P641.7 文献标志码: A 开放科学(资源服务)标识码(OSID):



地下水位动态变化是地下水资源储量变化的直观体现, 也是研究地下水赋存与演变特征的基础信息^[1], 获取地下水位数据是解决水资源问题的重要工作之一^[2]。实际工作中, 单井水位变化特征受区域补排特点和水文地质条件约束, 区域内各井点处水位随该区域地下水储量变化的规律不尽相同^[3-4]。地下水位变化情况目前已成为水利部、地方政府和相关单位开展地下水超采治理评估的主要依据^[5-7]。如何获取能够准确描述区域整体水位变化状况的代表值成为亟待解决的关键问题。本研究在总结常用方法的基础上, 使用 k -means 聚类分析对群井进行分类、泰森多边形去丛聚确定单井影响范围,

与传统水均衡法计算的水位变化值相结合建立群井观测水位数据和区域地下水储变量数据间的相关关系, 充分利用调查资料进行分析和计算并为群井分配合理的权重值, 达到在实际工作中能够利用权重值与各井新的观测水位数据快速计算区域代表性地下水位变化值的目的, 相对传统求解方法具有更高的简便性。

1 研究现状

前人研究工作中主要使用如下 4 种方法表征区域地下水位变化: 水位变化平均法、水位变化状况

收稿日期: 2022-01-14 修回日期: 2022-07-22 网络出版时间: 2022-07-26

网络出版地址: <https://kns.cnki.net/kcms/detail/13.1430.TV.20220726.1029.002.html>

基金项目: 国家自然科学基金面上项目(41972262); 河北自然科学基金优秀青年科学基金项目(D2020504032); 河北地质大学学生科研项目(KAZ202102)

作者简介: 康玮(1993—), 男, 河北石家庄人, 主要从事地下水科学方面研究。E-mail: kangwei_kevins@163.com

通信作者: 曹文庚(1985—), 男, 山东禹城人, 副研究员, 博士, 主要从事水文及水资源、地下水供水安全保障方面研究。

E-mail: caowengeng@mail.cgs.gov.cn

统计法、水均衡计算法和数值模拟法,在简便性和准确性上4种方法各有优劣。水位变化平均法和水位变化状况统计法两种方法直接基于统计数据进行分析,对含水层结构、补排条件和水位变化之间的关系阐述不足,工作量虽小、准确性差;水均衡计算法基于区域储变量算得代表性水位,考虑了计算

区内补排条件,准确性较高,但具有数据统计量大、种类多的缺点;数值模拟法通过建立模型模拟和刻画区域水位变化状况,在几种方法里准确性最高,缺点是建立模型需要大量参数支持和长时间的拟合验证工作。表1给出了4种方法的核心思想以及在实际应用过程中的优缺点。

表1 4种区域地下水位变化表征方法统计

Tab. 1 Characterization methods for groundwater level change in four regions

方法名称	核心思想	优点	缺点
水位变化平均法	将群井监测水位数据按含水层分类后加和平均代表区域地下水位的变化 ^[9-10]	具有简便性和直接性,广泛应用于生产单位 ^[11-12]	对水位变化分析评价不全面,科学性较差
水位变化状况统计法	统计群井监测水位数据,用群井发生相同水位变化的个数代表区域地下水位的变化趋势 ^[8]		
水均衡计算法	建立水均衡方程利用调查统计获取的源汇项数据间接求取研究区水位变化值 ^[13]	同时完成研究区水文地质条件分析和水位变化值求解,计算结果具有科学性	数据获取工作量大、种类多而精度不一,不便于直接快速获取结果 ^[14]
数值模拟法	建立研究区三维地下水数值模拟模型,模拟地下水流场在时间尺度上的变化过程,刻画研究区地下水流场的动态变化情况 ^[15-16]	全面直观地反映流场变化过程及各单井水位变化与整体水位变化之间的关系,计算结果精度高	需要大量数据支撑,建模工作持续时间较长,成本高,不适合快速获取研究区水位变化值

2 研究方法

本研究利用观测井连续观测数据,结合 k -means 聚类算法、泰森多边形去丛聚和水均衡法构建群井水位变化与区域储变量间的线性关系求解代表性地下水位变化值,其求解思路见图1。

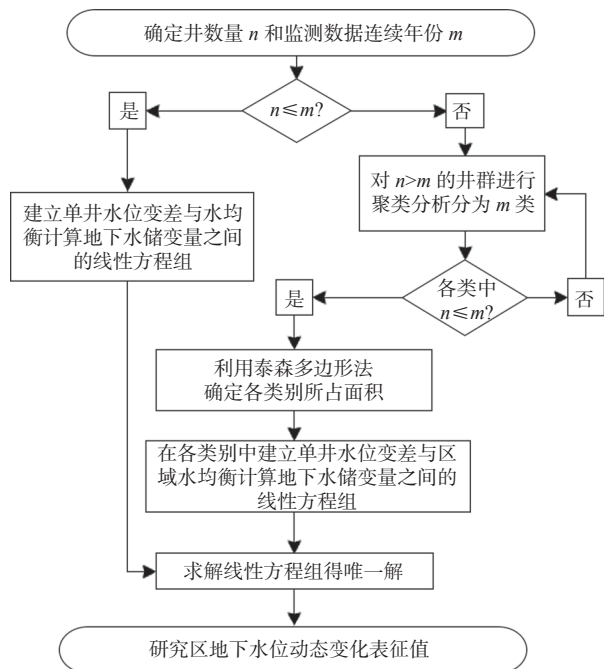


图1 研究思路

Fig. 1 Research flow chart

判断观测井群数量与水位监测数据连续周期数之间的关系。当观测井数量 n 小于或等于连续监测年数 m 时,可直接建立方程组进行求解。

实际工作中常见 n 大于 m 的情况,此时线性方程组中自变量数多于方程数,建立的方程组为超定方程,无法直接解得单井权重值。为解决这一问题,本次研究引入 k -means 聚类算法和泰森多边形去丛聚方法对单井和聚类后的虚拟观测井控制范围进行聚类分析,将单一方程多步求解确定唯一权重值,进而计算代表性地下水位变化值。其一般步骤为:

根据井数 n 和监测数据年数 m 利用 k -means 方法进行聚类,直至 $n \leq m$;

在各类别中利用泰森多边形方法确定单井控制面积;

建立以各分类中单井观测水位为自变量、以水均衡法计算水位值作为因变量的线性方程组;

汇总各井在各自分类中的权重值,计算单井在全区井群中的权重值;

根据单井分配权重值计算研究区代表性地下水位变化值。

以下具体解释在代表性地下水位变化值求解过程中的三个相互关联的重要方法。

2.1 线性方程组计算单井权重

对于分布在同一区域内的多个水文观测井,观测水位的变化是连续且相互关联的^[16],根据这一特点,已知 m 个连续年间 n 口井各自的水位变化值,可以建立观测值与计算值的线性方程组。建立的线性方程组用于描述群井观测水位变化值与水均

衡法计算的水位变化值之间的关系:

$$\begin{aligned} \alpha_{11}x_1 + \alpha_{21}x_2 + \alpha_{31}x_3 + \dots + \alpha_{n1}x_n &= \bar{X}_1 \\ \alpha_{12}x_1 + \alpha_{22}x_2 + \alpha_{32}x_3 + \dots + \alpha_{n2}x_n &= \bar{X}_2 \\ &\vdots \\ \alpha_{1m}x_1 + \alpha_{2m}x_2 + \alpha_{3m}x_3 + \dots + \alpha_{nm}x_n &= \bar{X}_m \end{aligned} \quad (1)$$

式中: α_{nm} 为第 n 口井 m 年的观测水位变化值, $m; x_n$ 为第 n 口井的单井权重值; \bar{X}_m 为第 m 年根据水均衡法算得的水位变化值, m 。

式(1)亦可矩阵化表述为

$$\begin{bmatrix} \alpha_{11} & \alpha_{21} & \dots & \alpha_{n1} \\ \alpha_{12} & \alpha_{22} & \dots & \alpha_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1m} & \alpha_{2m} & \dots & \alpha_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_m \end{bmatrix} \quad (2)$$

解上述方程组即可得到单井的权重值, 根据线性方程组解的规律(或行列式值的特征), 在去除等比例项后自变量个数与方程个数相等(即系数矩阵满秩)时方程有唯一解^[17-18]。此时便可获得单井的权重值, 利用该系数向量便可在后续工作中直接利用井群观测水位变化值计算出区域的代表性地下水变化值, 即:

$$\bar{X}' = \alpha'_1x_1 + \alpha'_2x_2 + \alpha'_3x_3 + \dots + \alpha'_nx_n \quad (3)$$

式中: α'_n 为第 n 口井的年水位变动观测值, $m; x_n$ 为第 n 口井的单井权重值; \bar{X}' 为计算年代表性地下水变化值, m 。

当观测井数 n 大于连续观察年数 m 时, 为保证方程组成立且有唯一解, 使用 k -means 聚类方法对井群进行分类降维后将一类井视为一个自变量构建方程组分步求解。

2.2 k -means 聚类方法

k -means 聚类算法是一种无监督学习方法, 是人工智能机器学习领域近年来常用的分类方法之一, 最早由 MacQueen 提出^[19]。其训练过程为: 根据多维度样本数据集中各样本间欧几里得距离分析样本之间的相似性, 在同条件下对样本进行判断, 将样本点与中心点欧几里得距离较近的数据聚为一簇, 每组数据有多个中心点, 从而可形成多个簇。每一个簇的样本相似度与其他簇内样本相比相似度较高, 经过多次训练, 最终可得到几个整体差异性较高的不同簇以实现聚类^[20], k -means 聚类算法相较于其他分类或聚类法具有客观性、科学性、广泛性等优点^[21], 聚类可用的维度(即参考特征的个数)相对灵活, 与其他数学方法相结合的研究效果较好^[22-23], 在气象科学^[24-26]、煤炭工业科学^[21, 27]、环境保

护与污染治理^[28-30] 以及水文地质学^[31] 等许多领域已有效果较好的应用实例。

本研究为解决由观测井数量多于连续年数据(即 $n > m$) 造成的方程组无解问题, 以井群观测值为基础数据, 利用 k -means 方法对井群进行聚类。在聚类后的 m 类井群中分别求得各类别中单井在该井群中的权重值, 再将每一类整体作为一个虚拟井, 类别数作为井数进行多次求解, 最终得到单井在全部井群中的权重值。

k -means 聚类算法具体实现步骤如下:

给出待聚类的多维数据集, 根据数据先验经验(如观测井数量和数据连续年)选择合适的聚类数值 k :

$$\begin{aligned} X &= \{x \mid x_{ij} \in R, i = 0, 1, 2, 3 \dots, n; j = q\}, \\ k &= 1, 2, 3, 4, \dots, m \end{aligned} \quad (4)$$

式中: x_{ij} 为多维向量; n 为样本总个数; q 为样本维度。

从数据集 X 中随机选取 K 个样本作为初始聚类中心, 分别计算其他所有样本点与确定的初始聚类中心间的欧氏距离^[32], 以最短距离作为判断标准将各样本分别聚入该簇中, 两点欧氏距离计算如下:

$$d = \sqrt{\sum_{j=1}^q (x_{ij} - \bar{x}_{kj})^2} \quad (5)$$

式中: \bar{x}_{kj} 为第 k 簇初始聚类中心。

聚类完成之后, 在新的每类数据中重新计算出新的聚类中心点, 再次判断新的聚类中心与该簇中其他样本间是否满足给定的限制: 若满足则聚类完成, 输出最终的聚类中心; 若不满足, 则将新的聚类中心作为初始, 再次进行聚类, 直到达到限制要求^[32-33]。通常, 限制条件通过规定误差平方和的值来规定, 误差平方和计算如下:

$$E = \sum_{i=1}^m \sum_{x \in C_i} d \quad (6)$$

式中: C_i 为第 i 簇样本。

在使用 k -means 聚类算法对井群进行聚类后, 需要利用泰森多边形法确定各单井的影响面积。

2.3 泰森多边形方法

为了构建用于支撑方程构建和权重值计算工作的均衡分区, 需要对研究区进行去丛聚处理, 确定各单井在井群中对应的影响范围。泰森多边形是一种广泛应用于定性分析、统计分析、邻近分析等研究的多边形去丛聚方法, 其最早由荷兰气候学家 A.H.Thiessen 提出并用于计算基于离散分布气象站记录的降雨量的地区平均降雨量^[34-36], 在去丛聚过

程中具有较高的精确度。使用泰森多边形去丛聚法对研究区面积进行分割,计算各观测井所在分区的面积,进而使用水均衡法计算各分区水位变化值。

泰森多边形去丛聚法主要步骤如下:

连接相邻观测井构成三角形网格;

对于每个独立的三角形网格,作其各边的中垂线,全部中垂线围成一个新的多边形网;

每一个独立多边形网格中包含唯一单井,此时该多边形网格所包围的区域即为该井对应的控制范围。

利用上述方法确定各单井的控制范围以及聚类后某个井群的总控制范围,使用水均衡法计算分区水位变化值,将结果用于权重值计算方程组。

3 算例分析

3.1 算例模型构建

使用 GMS 软件建立单元格大小为 70×70 单位长度,共约 7 000 个单元的理想情境算例模型。研究区渗透系数设定为 10 m/d ;给水度设定为 0.3 ;源汇项部分以降雨入渗量为主要补给项,排泄项主要为面状开采;应力期设定为 5 年。考虑实际工作中与算例区域面积大小相近的范围内井的分布情况,随机选择 5 个点位作为观测点,使用泰森多边形法

对研究区进行分割作为均衡分区见图 2。运行算例模型并记录算例区 5 年内每年年末时观测点水位数据,逐年作差获取观测点 4 组水位变差值,同时根据补给量和排泄量计算逐年地下水储变量并根据泰森多边形法分割的均衡区计算各均衡区水位的变化值,见表 2。

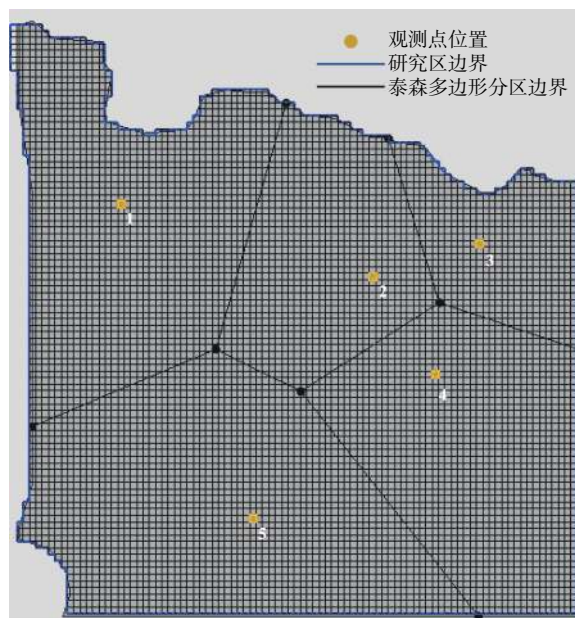


图 2 算例研究区概况

Fig. 2 Overview diagram of case study area

表 2 算例研究区模拟水位及计算水位变差

Tab. 2 Simulated water level and calculated water level variation in case study area

观测井编号	各应力期观测点水位变差/m				各应力期水均衡法计算水位变差/m			
	1	2	3	4	1	2	3	4
1	0.08	-0.68	0.31	-0.14				
2	-0.26	0.35	-0.40	0.30	0.59	0.76	0.79	0.83
3	-0.19	0.41	-0.32	0.34				
4	-0.36	0.33	-0.37	0.32	0.81	0.84	0.87	0.92
5	-0.44	0.25	-0.26	0.29	0.74	0.86	0.88	0.89

根据算例研究区概况和观测点数量建立线性方程组。算例中 $n=5, m=3$, 此时需要进行两次三元线性方程组求解的过程,使用 k -means 算法对观测点位置信息(西安 80 坐标系,大地坐标)进行聚类分析,算法生成的观测点位置信息和聚类结果见图 3。初次求解时将图 3(b)中第 II 类中 2、3、4 号 3 个观测点的数据代入方程组中进行求解;第二次求解时,将第 I 类中 1 号、第 II 类整体的以及第 III 类中的 5 号观测点数据代入方程求解。

3.2 算例结果

使用上述计算方法对算例数据进行计算,其中

前 3 组变差数据用于观测水位加权计算,第 4 组数据代入计算结果与实际值对比作为验证。代入求解后解得一组权重值见表 3。

利用观测点第 5 年水位变差数据和解得的水位变化权重值进行计算,得到第 5 年研究区整体水位变差为 0.23 m ;同时利用水均衡法计算第 5 年研究区整体水位变差为 0.21 m 。两者计算结果相差 0.02 m ,误差为 9.5% ,在算例中新方法的计算结果较为准确,说明该方法的在理想情境下的计算结果具有有效性。而当观测井数 n 大于连续观察年数 m 的区域时需对井群进行聚类和分步方程计算。

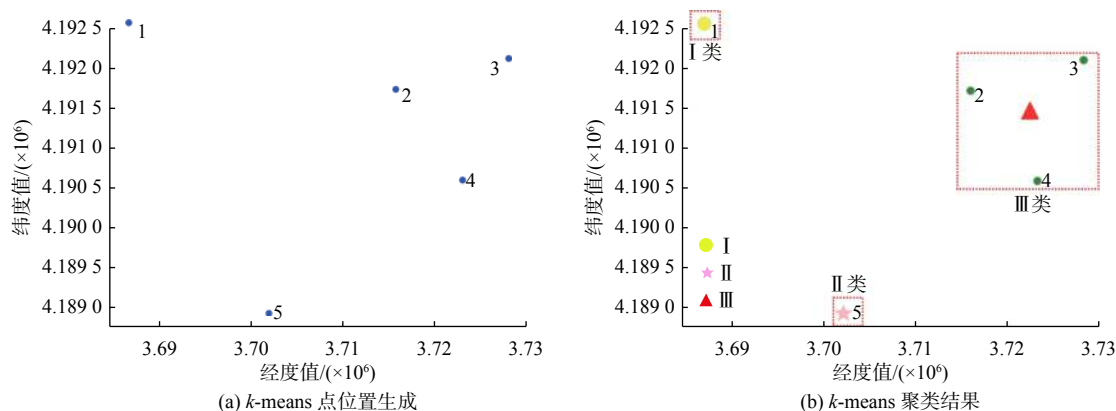


图 3 k-means 聚类算法观测点位置及聚类结果

Fig. 3 K-means clustering algorithm observation point location and clustering results

4 实例应用分析

为进一步验证方法的可行性和实用性,本次研究选取基础资料较为完整的定兴县行政单元作为实例研究区进行实例应用分析。

表 3 算例观测点水位变化权重值

Tab. 3 Weight value of water level change of observation well in example

观测点编号	1	2	3	4	5
权重值	0.360	0.186	0.301	0.080	0.354

4.1 定兴县模型构建

收集定兴县内共 8 口观测井 2015—2019 年连

表 4 2016—2019 年定兴县模拟水位及计算水位变差

Tab. 4 Simulated water level and calculated variation of water level in Dingxing County from 2016 to 2019

观测井编号	观测点水位变差/m				水均衡法计算水位变差/m			
	2016	2017	2018	2019	2016	2017	2018	2019
1	-0.92	-0.67	-0.11	-0.07				
2	-0.99	-0.87	-0.14	-0.16	-0.94	-0.76	-0.03	-0.11
3	-0.54	-0.48	-0.01	-0.20				
4	-0.83	-0.35	-0.13	0.31				
5	-1.00	-0.63	-0.06	0.28	-1.48	-0.72	-0.26	0.40
7	-0.56	-0.12	-0.23	0.17				
6	-0.66	-0.26	-0.33	-0.09				
8	-0.51	-0.30	-0.07	-0.04	-0.83	-0.49	-0.03	0.09

根据实例研究区概况和观测点数量建立线性方程组。此时 $n=8, m=3$, 需要进行三次三元线性方程组求解和一次二元线性方程组求解,使用 k-means 算法对观测点位置信息(西安 80 坐标系,大地坐标)进行聚类分析,聚类结果如图 5 所示,分别构建和求解第 I、II、III 类中观测井数据的线性方程组,再将 3 个类别视为 3 个虚拟观测井,利用计算水位变差值代入方程中求解各分类集合在整体中的权重

续 5 年年末时的观测水位数据用于计算 4 年的水位变差值,根据观测井在县区内的分布情况构建基于泰森多边形法分割的均衡分区,整理研究区内 2016—2019 年连续 4 年的各项源汇项数据。研究区内地下水自西北向东南流动,西北部边界为侧向流入边界,东南为人工流出边界,降雨入渗是主要的补水项,人工开采是主要排泄项,多数地区给水度分布在 0.08~0.10,靠近西北部边界处可达 0.18,渗透系数自西北山前 150 m/d 左右向东南逐渐下降到 20 m/d。根据水均衡法计算区域储变量,进而可算得各均衡区和研究区整体水位变差值。研究区连续年模拟水位变差及计算水位变差数据见表 4,研究区给水度分区和均衡分区见图 4。

值,最后利用各观测井在分类集合中的权重值和各分类在整体中的权重值算得各观测井的地下水动态变化权重值。

4.2 结果分析

使用上述计算方法对实例数据进行计算,其中前 4 年变差数据用于观测井权重值计算,第 5 年数据代入得到的权重值矩阵与实际值对比作为验证。代入求解后解得一组权重值见图 6 和表 5。

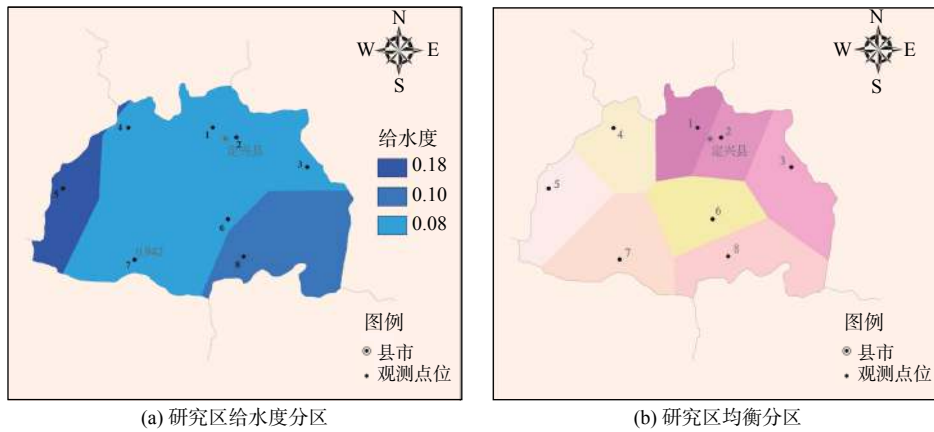


图4 定兴县研究区概况

Fig. 4 Overview map of research area in Dingxing County

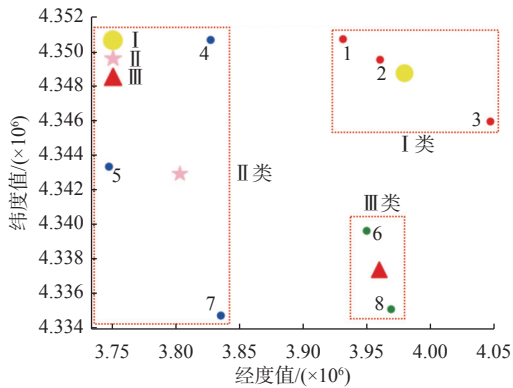


图5 研究区 *k-means* 聚类算法聚类结果

Fig. 5 *K-means* clustering algorithm clustering results in research area

由图6可知:位于研究区西北部的4号井和中部的6号井权重值均小于0.1,说明该点水位变差对整体区域水位变化的计算结果影响较小;东部的3号井和8号井权重值接近1.5,对计算结果的影响较大;西南部的5号井和7号井权重值则均在1左右。

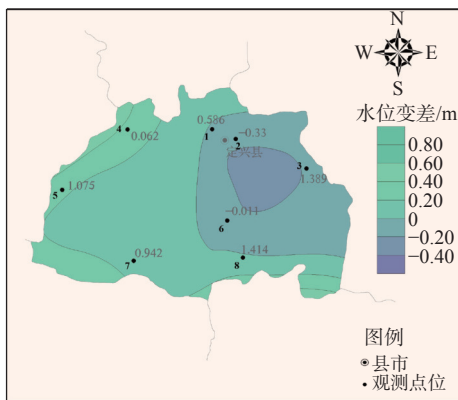


图6 研究区2019年水位变差及各点权重值

Fig. 6 Variation of water level and weighting values of each point in the study area in 2019

表5 实例观测井水位变化权重值

Tab. 5 Weight value of water level change in case observation well

观测井编号	1	2	3	4	5	6	7	8
权重值	0.568	-0.330	1.389	0.062	1.075	-0.011	0.942	1.414

聚类结果:1、2、3号井位于第I类内,3号和1号井处水位变化情况与分区整体水位动态呈正相关,2号井则呈负相关,根据权重值绝对值的大小分析得出3号井处水位变化对该分类整体水位变化的影响最大;4、5、7号井位于第II类内,其中4号井对该分类整体水位变化的影响较小,5号和7号井影响较大;6号和8号井属于第III类,该分类范围内整体水位变化主要受8号井处水位变化的影响,6号井处水位变化对整体几乎没有影响。

综合上述3个分区各单井获得的权重值,结合观测井第5年水位变差数据,利用公式(3)计算得到第5年研究区整体水位变差为0.16 m;而利用水均衡法计算水位值得到的第5年研究区整体水位变差为0.15 m。两者计算结果相差0.01 m,误差为6.7%,在实例中新方法的计算结果较为准确。在实验中发现误差的主要来源有两个:一是在计算区域水位变化时由水均衡法带来的计算误差,水均衡法先将区域概化为柱状的不规则四面体,根据柱状体内水量变化的体积与柱状体的横截面面积来计算水位变化值,在这一过程中将复杂的地质条件合并简化时会产生误差;另一个误差主要来源于监测井水位的获取,各井水位数据的测量误差对最终计算结果影响程度由该井的权重值决定。除此以外,井群的排布和密度对于计算结果亦有影响,在同一地区内增加或减少部分井或调整统计点位对于计算结果的影响程度会成为下一步重要的研究方向。

5 结论与讨论

本研究将 *k-means* 聚类分析、泰森多边形去丛聚和传统的水均衡算法相结合,以实际水位监测数据为基础提出一套计算代表性地下水位变化值

的系统方法,通过算例和实例对方法的准确性和实用性进行了验证,算例研究区中该方法较水均衡法计算水位变化值误差为9.5%,在定兴县实例研究区这一误差结果为6.7%,结果均较为准确,同时在获取结果的过程中考虑了水文地质条件和地下水位变差,相较于水位变化平均法和水位变化状况统计法新方法具有更高的合理性。

在实例研究区定兴县,对8口观测井的4年连续水位变化数据和储变量数据进行计算,得到一组系数:0.568、-0.330、1.389、0.062、1.075、-0.011、0.942、1.414。使用这一组系数与末年各井水位变差值算得的定兴县2019年区域水位变化为0.16 m,这与传统的水均衡法算得的结果仅相差0.01 m,计算结果准确度很高且数据统计工作量可极大减少。

该方法适用于连续年监测数据较少但观测井数较多的地区,在井密度较小和井的分布极不均匀的地区需根据实际情况增设观测井位以确保结果的准确性。获取单井权重值后,可在之后的统计中直接利用各井观测水位变化值与对应系数算得研究区整体代表性地下水位变化值,据此对区域水位整体变化进行评价,极大地减少了统计和评估的工作量;同时,采用多种科学方法结合建立的方程描述了具有不同地质条件的各井点处水位变化与区域整体水位变化之间的相关性关系,与传统方法相比能够给出更加准确和显著的超采治理工作效果,在实际工作中具有较高的实用价值。

参考文献(References):

- [1] 胡立堂,郭建丽,张寿全,等.永定河生态补水的地下水水位动态响应[J]. *水文地质工程地质*, 2020, 47(5): 5-11. (HUL T, GUO J L, ZHANG S Q, et al. Dynamic response of groundwater table to ecological recharge of Yongding River[J]. *Hydrogeology & Engineering Geology*, 2020, 47(5): 5-11. (in Chinese)) DOI: 10.16030/j.cnki.issn.1000-3665.202008027.
- [2] 束龙仓,王哲,袁亚杰,等.近40年三江平原典型区土地利用变化及其对地下水的影响[J]. *水利学报*, 2021, 52(8): 896-906. (SHU L C, WANG Z, YUAN Y J, et al. Land use change and its influence on groundwater in typical area of Sanjiang Plain in recent 40 years[J]. *Journal of Hydraulic Engineering*, 2021, 52(8): 896-906. (in Chinese)) DOI: 10.13243/j.cnki.slxb.20210284.
- [3] 凤蔚,王晓燕,刘振英,等.松散岩类承压含水层影响半径计算方法[J]. *人民黄河*, 2017, 39(12): 57-61. (FENG W, WANG X Y, LIU Z Y, et al. Calculation method of influence radius of confined aquifer in loose rock[J]. *Yellow River*, 2017, 39(12): 57-61. (in Chinese)) DOI: 10.3969/j.issn.1000-1379.2017.12.014.
- [4] 卢放,刘伟朋,胡文广,等.抽水试验影响半径的 α 读数与自然电位特征[J]. *南水北调与水利科技*, 2019, 17(6): 163-169. (LU F, LIU W P, HU W G, et al. Influence radius of pumping test α readings and natural potential characteristics[J]. *South-to-North Water Transfers and Water Science & Technology*, 2019, 17(6): 163-169. (in Chinese)) DOI: 10.13476/j.cnki.ns-bdqk.2019.0147.
- [5] 河北省水利厅水资源处.河北省地下水超采综合治理试点第三方评估结果公布[J]. *河北水利*, 2018(11): 5. (Department of Water Resources Hebei Province. Third party evaluation results for pilot work of comprehensive treatment of groundwater overexploitation in Hebei Province were published[J]. *Hebei Water Resources*, 2018(11): 5. (in Chinese))
- [6] 杨得瑞.华北地下水超采综合治理河湖地下水回补试点探索与效果评估[J]. *中国水利*, 2020(13): 15-16. (YANG D R. Pilot exploration and effect evaluation on comprehensive treatment of groundwater overexploitation in north China for groundwater recharge in rivers and lakes[J]. *China Water Resources*, 2020(13): 15-16. (in Chinese)) DOI: 10.3969/j.issn.1000-1123.2020.13.013.
- [7] 曹文庚,杨会峰,高媛媛,等.南水北调中线受水区保定平原地下水质量演变预测研究[J]. *水利学报*, 2020, 51(8): 924-935. (CAO W G, YANG H F, GAO Y Y, et al. Study on groundwater quality evolution forecast of Baoding Plain in water receiving area of middle route of South-to-North Water Transfer[J]. *Journal of Hydraulic Engineering*, 2020, 51(8): 924-935. (in Chinese)) DOI: 10.13243/j.cnki.slxb.20200035.
- [8] 吕晨旭,贾绍凤,季志恒.近30年来白洋淀流域平原区地下水水位动态变化及原因分析[J]. *南水北调与水利科技*, 2010, 8(1): 65-68. (LYU C X, JIA S F, JI Z H. Dynamic change and cause analysis of groundwater level in plain area of Baiyangdian watershed in recent 30 years[J]. *South-to-North Water Transfers and Water Science & Technology*, 2010, 8(1): 65-68. (in Chinese)) DOI: 10.3969/j.issn.1672-1683.2010.01.017.
- [9] 杨亮平,姜振蛟,赵宜婷,等.呼和浩特市地下水水位动态变化及趋势预测研究[J]. *水文地质工程地质*, 2009, 36(4): 46-49. (YANG L P, JIANG Z J, ZHAO Y T, et al. Dynamic change and trend forecast of groundwater level in Hohhot[J]. *Hydrogeology & Engineering Geology*, 2009, 36(4): 46-49. (in Chinese))

- DOI: [10.3969/j.issn.1000-3665.2009.04.011](https://doi.org/10.3969/j.issn.1000-3665.2009.04.011).
- [10] 郑新奇, 张乃兴, 李新运. 莱州湾东南沿岸地下水水位动态与海水入侵相关规律研究[J]. *水文地质工程地质*, 1997(3): 6-9. (ZHENG X Q, ZHANG N X, LI X Y. Study on the correlation between groundwater level dynamics and seawater intrusion along the southeast coast of Laizhou Bay[J]. *Hydrogeology & Engineering Geology*, 1997(3): 6-9. (in Chinese)) DOI: [10.16030/j.cnki.issn.1000-3665.1997.03.002](https://doi.org/10.16030/j.cnki.issn.1000-3665.1997.03.002).
- [11] 张丽, 田晓龙. 巩义市城区地下水动态特征分析[J]. *华北水利水电大学学报(自然科学版)*, 2015, 36(3): 24-27. (ZHANG L, TIAN X L. Analysis of groundwater dynamic characteristics in Gongyi City[J]. *Journal of North China University of Water Resources and Electric Power(Natural Science Edition)*, 2015, 36(3): 24-27. (in Chinese)) DOI: [10.3969/j.issn.1002-5634.2015.03.006](https://doi.org/10.3969/j.issn.1002-5634.2015.03.006).
- [12] 李晓芳, 闫琴, 张少博, 等. 新疆玛纳斯河灌区地下水特征研究[J]. *中国农村水利水电*, 2017(3): 213-216. (LI X F, YAN Q, ZHANG S B, et al. Study on groundwater characteristics of Manas River irrigation area in Xinjiang[J]. *China Rural Water and Hydropower*, 2017(3): 213-216. (in Chinese)) DOI: [10.3969/j.issn.1007-2284.2017.03.046](https://doi.org/10.3969/j.issn.1007-2284.2017.03.046).
- [13] 乔晓英, 肖平, 王继华, 等. 郑州地下水均衡试验场的改建工程: 试验场监测资料的推广应用核心问题[J]. *水文地质工程地质*, 2020, 47(1): 46-52. (QIAO X Y, XIAO P, WANG J H, et al. Reconstruction project of Zhengzhou groundwater balance test site: The core issue of popularization and application of monitoring data of test site[J]. *Hydrogeology & Engineering Geology*, 2020, 47(1): 46-52. (in Chinese)) DOI: [10.16030/j.cnki.issn.1000-3665.201908001](https://doi.org/10.16030/j.cnki.issn.1000-3665.201908001).
- [14] 束龙仓, 张曼琦, 李虎, 等. 两种典型边界条件下的趵突泉泉域水均衡分析[J]. *南水北调与水利科技(中英文)*, 2021, 19(3): 427-436, 445. (SHU L C, ZHANG M Q, LI H, et al. Analysis of water balance in Baotu Spring basin under two typical boundary conditions[J]. *South-to-North Water Transfers and Water Science & Technology*, 2021, 19(3): 427-436, 445. (in Chinese)) DOI: [10.13476/j.cnki.nsbqk.2021.0046](https://doi.org/10.13476/j.cnki.nsbqk.2021.0046).
- [15] 张将伟, 卢文喜, 曲延光, 等. 基于Monte Carlo方法的地表水地下水耦合模拟模型不确定分析[J]. *水利学报*, 2018, 49(10): 1254-1264. (ZHANG J W, LU W X, QU Y G, et al. Uncertainty analysis of surface water and groundwater coupling simulation model based on Monte Carlo method[J]. *Journal of Hydraulic Engineering*, 2018, 49(10): 1254-1264. (in Chinese)) DOI: [10.13243/j.cnki.slxb.20180402](https://doi.org/10.13243/j.cnki.slxb.20180402).
- [16] 李肖男, 张红武, 钟德钰, 等. 黄河下游河道治理三维数值模拟研究[J]. *水利学报*, 2017, 48(11): 1280-1292. (LI X N, ZHANG H W, ZHONG D Y, et al. Three-dimensional numerical simulation study on river regulation in the lower reaches of the Yellow River[J]. *Journal of Hydraulic Engineering*, 2017, 48(11): 1280-1292. (in Chinese)) DOI: [10.13243/j.cnki.slxb.20160958](https://doi.org/10.13243/j.cnki.slxb.20160958).
- [17] 陈玉文, 嵇绍春, 钱树华, 等. 线性代数[M]. 南京: 南京大学出版社, 2019. (CHEN Y W, JI S C, QIAN S H, et al. Linear algebra[M]. Nanjing: Nanjing University Press, 2019. (in Chinese))
- [18] 高翔宇. 行最简形矩阵唯一性证明及其应用[J]. *高等数学研究*, 2020, 23(5): 10-12. (GAO X Y. Uniqueness proof of line minimum matrix and its application[J]. *Studies in College Mathematics*, 2020, 23(5): 10-12. (in Chinese)) DOI: [10.3969/j.issn.1008-1399.2020.05.005](https://doi.org/10.3969/j.issn.1008-1399.2020.05.005).
- [19] 庄承彬, 陈晓宏, 黄薇颖, 等. 一种基于三维指标因子的流域多年径流丰枯k-means聚类法[J]. *生态环境学报*, 2010, 19(6): 1354-1357. (ZHUANG C B, CHEN X H, HUANG W Y, et al. A k-means clustering method for multi-year runoff of watershed based on three-dimensional index factors[J]. *Ecology and Environmental Sciences*, 2010, 19(6): 1354-1357. (in Chinese)) DOI: [10.3969/j.issn.1674-5906.2010.06.017](https://doi.org/10.3969/j.issn.1674-5906.2010.06.017).
- [20] 林涛, 赵璨. 最近邻优化的k-means聚类算法[J]. *计算机科学*, 2019, 46(S2): 216-219. (LIN T, ZHAO C. K-means clustering algorithm with nearest neighbor optimization[J]. *Computer Science*, 2019, 46(S2): 216-219. (in Chinese))
- [21] 张嘉睿, 夏玉成, 李涛, 等. Entropy-Kmeans方法在煤层气开发前景评价中的应用[J]. *煤矿安全*, 2020, 51(8): 158-163. (ZHANG J R, XIA Y C, LI T, et al. Application of Entropy-Kmeans method in coal-bed gas development prospect evaluation[J]. *Safety in Coal Mines*, 2020, 51(8): 158-163. (in Chinese)) DOI: [10.13347/j.cnki.mkaq.2020.08.035](https://doi.org/10.13347/j.cnki.mkaq.2020.08.035).
- [22] 秦悦, 丁世飞. 半监督聚类综述[J]. *计算机科学*, 2019, 46(9): 15-21. (QIN Y, DING S F. Semi-supervised clustering overview[J]. *Computer Science*, 2019, 46(9): 15-21. (in Chinese)) DOI: [10.11896/j.issn.1002-137X.2019.09.002](https://doi.org/10.11896/j.issn.1002-137X.2019.09.002).
- [23] 廉文娟, 史丹丹, 安其立, 等. 基于最大期望的初始聚类中心选择算法[J]. *软件*, 2020, 41(5): 1-5. (LIAN W J, SHI D D, AN Q L, et al. Initial clustering center selection algorithm based on maximum expectation[J]. *Software*, 2020, 41(5): 1-5. (in Chinese))

- nese)) DOI: [10.3969/j.issn.1003-6970.2020.05.001](https://doi.org/10.3969/j.issn.1003-6970.2020.05.001).
- [24] 毛开银, 赵长名, 何嘉. 基于XGBoost的10 m风速订正研究[J]. *成都信息工程大学学报*, 2020, 35(6): 604-609. (MAO K Y, ZHAO C M, HE J. 10 m wind speed correction study based on XGBoost[J]. *Journal of Chengdu University of Information Engineering*, 2020, 35(6): 604-609. (in Chinese)) DOI: [10.16836/j.cnki.jcui.2020.06.004](https://doi.org/10.16836/j.cnki.jcui.2020.06.004).
- [25] NGOC H L, HUYEN T, PHI N, et al. MCCVA: A new approach using SVM and kmeans for load balancing on cloud[J]. *International Journal on Cloud Computing: Services and Architecture*, 2020, 10(3): 1-14. DOI: [10.5121/ijccsa.2020.10301](https://doi.org/10.5121/ijccsa.2020.10301).
- [26] MOURAD L, KARIM L, JEAN M B, et al. Improving satellite rainfall estimation from MSG data in Northern Algeria by using a multi-classifier model based on machine learning[J]. *Journal of Hydrology*, 2020, 584: 124705. DOI: [10.1016/j.jhydrol.2020.124705](https://doi.org/10.1016/j.jhydrol.2020.124705).
- [27] 谢玮, 毕臣臣, 刘学清, 等. 基于PCA-Kmeans++的煤层气多属性融合聚类分析方法研究[J]. *煤炭技术*, 2019, 38(5): 53-56. (XIE W, BI C C, LIU X Q, et al. Research on multi-attribute fusion cluster analysis method of coal-bed gas based on PCA-Kmeans++[J]. *Coal Technology*, 2019, 38(5): 53-56. (in Chinese)) DOI: [10.13301/j.cnki.ct.2019.05.018](https://doi.org/10.13301/j.cnki.ct.2019.05.018).
- [28] 吴苏保, 王慧锋, 颜秉勇, 等. 基于ViBe模型的VOCs泄漏自动识别算法[J]. *控制工程*, 2020, 27(11): 1970-1974. (WU S B, WANG H F, YAN B Y, et al. Automatic leak identification algorithm for VOCs based on ViBe model[J]. *Control Engineering*, 2020, 27(11): 1970-1974. (in Chinese)) DOI: [10.14107/j.cnki.kzgc.20180546](https://doi.org/10.14107/j.cnki.kzgc.20180546).
- [29] FRANCIS K M, ZHE G, MATEETE B, et al. Sustainable recommendation domains for scaling agricultural technologies in Tanzania[J]. *Land Use Policy*, 2017, 66: 34-48. DOI: [10.1016/j.landusepol.2017.04.028](https://doi.org/10.1016/j.landusepol.2017.04.028).
- [30] 黄天天, 张麒, 彭佳红. 农业面源污染聚类分析[J]. *电脑知识与技术*, 2014, 10(25): 5988-5990. (HUANG T T, ZHANG Q, PENG J H. Cluster analysis of agricultural non-point source pollution[J]. *Computer Knowledge and Technology*, 2014, 10(25): 5988-5990. (in Chinese)) DOI: [10.14004/j.cnki.ckt.2014.0286](https://doi.org/10.14004/j.cnki.ckt.2014.0286).
- [31] ADNAN R M, KHOSRAVINIA P, KARIMI B, et al. Prediction of hydraulics performance in drain envelopes using Kmeans based multivariate adaptive regression spline[J]. *Applied Soft Computing Journal*, 2021, 100. DOI: [10.1016/J.ASOC.2020.107008](https://doi.org/10.1016/J.ASOC.2020.107008).
- [32] 王琢, 荀亚玲, 张继福. 一种基于K-means的关联规则聚类算法[J]. *太原科技大学学报*, 2016, 37(6): 429-437. (WANG Z, XUN Y L, ZHANG J F. K-means-based clustering algorithm for association rules[J]. *Journal of Taiyuan University of Science and Technology*, 2016, 37(6): 429-437. (in Chinese)) DOI: [10.3969/j.issn.1673-2057.2016.06.003](https://doi.org/10.3969/j.issn.1673-2057.2016.06.003).
- [33] JABI M, PEDERSOLI M, MITICHE A, et al. Deep clustering: On the link between discriminative models and k-means[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(6): 1887-1896. DOI: [10.1109/TPAMI.2019.2962683](https://doi.org/10.1109/TPAMI.2019.2962683).
- [34] 李少华, 刘远刚, 王延忠. 泰森多边形在地质数据去丛聚中的应用[J]. *物探与化探*, 2011, 35(4): 562-564. (LI S H, LIU Y G, WANG Y Z. Application of Tyson polygon in de-clustering of geological data[J]. *Geophysical and Geochemical Exploration*, 2011, 35(4): 562-564. (in Chinese)) DOI: [CNKI:SUN:WTYH.0.2011-04-029](https://doi.org/CNKI:SUN:WTYH.0.2011-04-029).
- [35] SHIN J C. Rain gauge significance evaluation based on mean hyetographs[J]. *Natural Hazards*, 2011, 56(3): 767-784. DOI: [10.1007/s11069-010-9589-3](https://doi.org/10.1007/s11069-010-9589-3).
- [36] SEOK H H, KUE B K, DAWSEI H. Comparison of methods to estimate areal means of short duration rainfalls in small catchments, using rain gauge and radar data[J]. *Journal of Hydrology*, 2020, 588: 1-4. DOI: [10.1016/j.jhydrol.2020.125084](https://doi.org/10.1016/j.jhydrol.2020.125084).

Calculation method of representative groundwater level based on k-means cluster analysis and Thiessen de-clustering

KANG Wei^{1,2}, CAO Wengeng^{1,3}, XU Lixia⁴, NAN Tian^{1,3}, GAO Yuanyuan⁵, NIE Ziyi²

(1. The Institute of Hydrogeology and Environmental Geology, CAGS, Shijiazhuang 050061, China; 2. Hebei GEO University, Shijiazhuang 050031, China; 3. National Observation and Research Station on Groundwater and Land Subsidence in Beijing-Tianjin-Hebei Plain, Shijiazhuang 050061, China; 4. Laiyuan County Water Resources Bureau, Baoding 074099, China; 5. Bureau of South to North Water Transfer of Planning, Designing and Management, Ministry of Water Resources, Beijing 100038, China)

Abstract: The dynamic change of groundwater level is an intuitive reflection of groundwater reserve change and

basic information for studying the characteristics of groundwater occurrence and evolution. The variation characteristics of water level in a single well are affected by supplementary drainage characteristics and hydrogeological conditions. The water level of distributed wells at different points may have different trends with total groundwater reserve changes in a region. How to obtain representative values which can describe the overall regional water level changes becomes a key problem to be solved urgently.

A correlation between the observed water level of multiple wells and the water level calculated by regional groundwater reservoir variables was constructed based on water balance method. The k -means clustering analysis method and Thiessen polygon were used to solve the multi-solution problem caused by the gap between the number of wells and the length of continuous water level observations of a single well. By solving the linear equation, a set of weighting coefficients that can describe the relationship between a single well water level change and the whole area water level change was obtained. These weighting coefficients and water level changes of corresponding wells were used to quickly calculate regional representative groundwater level changes.

A new method was calculated and validated by numerical simulation and regional example. The results show that in an ideal example, using the water level variation data of five selected observation points and corresponding weight values, the annual variation of overall water level was 0.23 m. The calculated result was 0.02 m, different from the verified annual water level variation of 0.21 m calculated by the water balance method. In an ideal situation, the new method showed good accuracy and it reached 90.5%. In a typical area, a set of coefficients were obtained by water level data of 8 wells in Dingxing County for 5 consecutive years and groundwater storage variable data. The representative groundwater level change in Dingxing County in 2019 was calculated as 0.16 m by an obtained coefficient. This result was 0.01 m in difference from that of the water balance method, with an accuracy of 93.3%.

Based on actual water level monitoring data, a systematic method for calculating representative groundwater level changes was proposed in the new method by combining various methods. The accuracy and practicability of the method were verified by ideal examples and in Dingxing County. In the ideal example, the error of this method was 9.5% compared with that of the water balance method. This error was 6.7% in the calculation result taking Dingxing County as an example. New method can be used to calculate regional representative groundwater level changes. During the experiment, it was found that the new method need to add observation wells to ensure the accuracy of results according to the actual situation in areas with less monitoring data but more observation wells or less well distribution density or uneven distribution. After obtaining a set of weighting coefficients of a single well, the observed water level change values and corresponding coefficients of each well can be directly used to calculate the groundwater level change values of the whole representative groundwater level in the study area. The overall change in regional water level can be evaluated, which greatly reduces the workload of statistics and evaluation.

Key words: groundwater dynamics; k -means clustering; Thiessen polygon; representative water level; evaluation of over exploitation control