

应用信息熵方法对区域地下水观测网的优化研究

陈植华, 丁国平

(中国地质大学工程学院, 湖北武汉 430074)

摘要: 数据冗余是观测网优化需要解决的主要问题之一, 冗余数据既造成数据噪音又增加观测网运行成本. 减少数据冗余的主要手段是减少观测孔的数量, 但前提是不影响原有观测网提供信息的能力. 作者基于信息熵概念和随机技术的结合, 提出了一种优化观测孔数量的方法. 这种方法的基本原理是利用信息熵理论来评价每一个观测孔数据信息含量大小, 计算观测网中每一对观测孔之间的信息流交换, 并建立观测孔之间信息传递与距离的统计关系, 这种关系是判断冗余性观测孔的基础. 这种方法被用来优化河北平原区域地下水观测网, 不仅解决了数量问题, 同时也可确定具体冗余性观测孔. 最后结果对比证明, 将河北平原地下水观测孔的数量减少 26%, 几乎没有影响现有观测网提供数据信息的能力.

关键词: 观测网; 冗余数据; 优化; 信息熵; 信息传递.

中图分类号: P641.74

文献标识码: A

文章编号: 1000-2383(2001)05-0517-07

作者简介: (1956-) 男, 副教授, 在职博士, 主要研究方向是 GIS 与随机方法在水资源与地质环境的评价与模拟方面的应用.

1 地下水观测网的数据冗余问题

地下水观测网优化的目的之一是解决观测网的冗余数据问题. 所谓冗余数据是指这部分数据不包含有价值的信息, 去除这些数据后不影响或并不明显影响观测网提供信息的能力. 观测网中数据冗余性表现在两个不同的部分, 一是因某些观测孔空间位置过密, 导致部分观测数据冗余, 这部分观测孔属于冗余性观测孔; 另一方面是因为过短的观测时间或过密的观测频率产生的数据冗余, 这些数据不能更多地提供地下水要素变化的有关信息, 因此也属于冗余性成分. 观测网优化就是寻找某种方法来确定这些冗余性成分的存在, 并在保证不影响原来观测网所具有的提供信息能力的前提下消除这些冗余性成分, 这就是观测网的优化问题. 本文研究的主要内容是解决观测孔数量的冗余问题.

观测孔数量的冗余在我国目前的地下水观测网中普遍存在, 问题的主要原因可以从两个方面寻找答案. 一是观测网建设主要始于 20 世纪 60 70 年代, 限于当时理论和技术发展水平、限于对地下水空间

分布和运移规律的理解, 观测孔的空间布局主要依据水文地质分析和地形地貌条件等定性方法来确定, 这些观测孔的空间分布是否合理主要取决于研究者对研究区水文地质条件的认识和经验的积累, 因此经验性成分较多; 第二个原因表现在组成观测网的观测孔来源成分复杂, 受经济成本约束, 许多观测孔来源于一些地质、水文地质勘探孔, 地下水开采孔或其他目的施工所保留的钻孔, 因此, 开采利用地下水集中的地段观测孔的密度“自然”就大, 而偏僻的山区以及地下水开采条件不佳的地段观测孔的密度“自然”就稀, 这种“自然形成”的地下水观测网不可避免地具有许多冗余性成分.

数据冗余的直接弊端导致经济投入与信息产出比不匹配. 面对市场经济愈来愈占主导地位的今天, 如何以最经济的成本来保证地下水观测网的正常运行, 以最小的投入来获取必要的信息数据等是我们已经面对的现实; 另一个没有引起重视的问题是冗余性数据带来的噪音成分, 这种噪音增加了数据分辨的困难, 使人们对现有数据和观测网的可靠性产生模糊的认识, 仅从观测站、点的数量多少来评价观测网的精度和可靠性. 因此, 面临现代社会对观测网的要求, 我们已经难以回避观测网的数据冗余问题.

收稿日期: 2001-03-15

基金项目: 国家计委防灾、减灾项目: 计国地函[1998]29号.

降低观测网的运行成本,尽可能减少数据的冗余性,最主要的方法就是减少观测孔(站的数量),使观测孔数量合理化,当然这必须在保证现有观测网收集数据信息的能力没有降低或明显降低的约束条件下才可行。

国内外对观测孔数量的优化研究已有部分报道,Morin 等^[1]在 1979 年首先应用统计技术中的主因子分析法研究降水观测网站的合理化问题,1986 年 Galeati 等^[2]将主因子分析法与聚类分析相结合,进一步研究了降水观测网的合理化问题。1991 年,加拿大的 Burn 等^[3]采用层次聚类技术研究了加拿大 Pembina 流域盆地水文观测站的优化选择。这些研究方法对观测网优化的技术思路,主要是根据某种统计指标(相似性指标)将那些具有统计相似的观测站分为一类,结合研究者或决策者的判断在不同的分类结果中选择具有代表性的观测站,同时除去部分站点以达到减少观测站数量的目的。这种方法的优点是简单,但在确定冗余性观测孔时主要还是依据人为的判断。1990 年美国地调局的 Timothy 等^[4]利用统计方法进行地下水水质观测网数量优化研究,从估计精度与样本数量大小之间存在的统计关系出发,在假设置信水平为 95 %和同一标准差的前提下从原有的 120 个观测点减少到 99 个。这种研究方法立足于统计基础,从不同样本(即观测站的数量)与统计精度的关系来确定合理的观测孔数量。然而,这只能解决所需的观测孔统计数量,无法确定观测网中具体哪些属于冗余观测孔。仅有一个统计数据标准和合适的统计样本数量是无法解决观测网具体的冗余性问题的,也就是说统计方法的缺陷是一种对整个观测网所有观测孔的综合评价,它无法具体到每一观测孔的选择和取舍。

作者基于信息熵概念和随机理论的结合,提出了一种优化观测孔数量的方法。这种方法的基本原理是利用信息熵理论和技术来评价每一个观测孔数据的信息含量大小,是对每一个观测孔提供信息的能力的评价,观测网可视为信号通讯系统,计算每一对观测孔之间的信息传递大小,这种信息联系程度作为判断空间上观测孔冗余性的基础。这种方法可以具体确定哪些观测孔需要保留,哪些属于冗余性的观测孔可以去除而不影响到观测网提供数据信息的能力。这种方法应用于河北平原区域地下水观测网的优化研究,最后的对比结果证明,即使将目前观测孔的数量减少 26 %后,对现有观测网提供信息的

能力几乎没有影响。

2 信息熵优化地下水观测网的原理

2.1 观测网系统与信号通讯系统

地下水观测网系统可以看成是一个以地下水流场或地下水化学场为载体的信号通讯网络,各观测孔均可看成是地下水要素动态变化信号的接受者,各观测孔之间存在信息交流或存在信息传递。这种“通讯网络”具备信号传输的基本特征:(1)信号的可传递性:同一地下水流动系统内任一点水动力场的波动将在所有观测点中反映,尽管这种反映的程度和时间可能出现差异,但表现出同一流场具有密切的信息或信号传递;(2)信号传递过程的衰减性:在统一的地下水流场,观测孔之间的信号交流随空间距离的增加而减弱,即信号传递过程中出现一定的衰减性;(3)信号之间的相关性:不同层次地下水流动系统之间的水动力和水化学联系必然出现差异,表现为彼此信号传递和信息转换密切程度的差异。正是由于观测网系统具有通讯网络相似的上述一些特征,我们可以借用信息熵理论来对观测网系统信息传递能力进行评价和优化。

2.2 信息熵方法的基本原理

信息熵是信号通讯理论中用来评价随机信号中所含信息量大小的基本概念之一,它对随机信号中所含信息量的度量依据随机信号出现的概率大小。如果随机信号出现的概率越大,则因确定性成分大而信息量越小,如果是确定性事件则信息量等于零。只有我们收到的随机信号能够帮助减少事件的非确定性时,这表示我们获得了信息。因此,不确定性与信息熵通常是同一含义。类似地也可以理解地下水观测孔信息熵的物理意义:某观测孔具有的系列观测数据即被认为是一随机信号,如果地下水位处于稳定状态,这是一种确定性事件,观测孔数据没有提供新信息,因此其信息熵等于零;如果该点水位(或水质)变化大,数据(随机信号)分布在不同的概率区间,此时观测数据包含有新的信息。观测孔监测的要素变化愈大,表明该点水位或水质的非确定性愈大,信息熵也越大,观察孔也愈有价值。因此,可以基于观测孔数据的信息熵大小来评价观测孔提供信息的能力。信息熵 $H(X)$ 大小的计算可根据下式^[5]:

$$H(X) = - \sum_{i=1}^N p(x_i) \ln p(x_i), \quad (1)$$

式中, $p(x_i)$ 是随机信号 X (水位或水质变化 x_i) 发生的概率 ($x_i, i = 1, 2, 3, \dots, N = \text{数据系列长度}$), 信息熵 $H(X)$ 的单位取决于对数形式, 自然对数时 $H(X)$ 的单位是奈比(Napier)。

定量刻画观测孔之间信息联系程度大小需要理解信息熵理论中的互信息 (mutual information) 概念。互信息 $T(X_i, X_j)$ 的物理意义在这里表示两个随机信号 (即观测孔对) X_i, X_j 之间信息传递量的大小, $T(X_i, X_j)$ 也称事件 X_i 给事件 X_j 的信息传递量。 $T(X_i, X_j)$ 可根据信息熵概念和公式推导出基本的计算方法^[6] :

$$T(X_i, X_j) = H(X_i) - H(X_i | X_j), \quad (2)$$

或

$$T(X_i, X_j) = H(X_i) + H(X_j) - H(X_i, X_j), \quad (3)$$

其中

$$H(X_i | X_j) = - \sum_{i=1}^N \sum_{j=1}^N p(x_i, x_j) \ln p(x_i | x_j) dx_i dx_j. \quad (4)$$

式中, $p(x_i, x_j), p(x_i | x_j)$ 分别为 X_i, X_j 的联合概率和条件概率; $H(X_i | X_j)$ 称条件熵, 物理意义可以理解为随机信号 X_i, X_j 之间信息传递过程中的信息损失量; $H(X_i, X_j)$ 是联合熵。

如果随机信号的概率分布属于正态或对数正态分布形态, Harmancioglu 等^[6] 和 Ozkul^[7] 分别在 1992, 2000 年导出了正态分布情况下多随机变量联合熵的通用计算公式:

$$H(X) = (M/2) \ln 2 + (1/2) \ln |C| + (M/2) - \ln(x). \quad (5)$$

式中, M 为变量的个数, C 为多变量协方差矩阵, x 是多变量分类的间隔大小, 在这里要求多变量的 x 为一常数。该公式可以适应单、双和多变量不同情况下信息熵和联合熵的计算^[6,7]。

互信息之所以可以用来定量表示观测孔之间的信息联系大小是由于互信息具有几个基本性质: (1) 当观测孔 X_i, X_j 之间互为独立时, 由于 $H(X_i) + H(X_j) = H(X_i, X_j)$; $T(X_i, X_j) = 0$, 表示两个互为独立的观测孔之间没有联系, 不存在信息传递或信息转换; (2) 当观测孔 X_i, X_j 之间具有统计联系时, 由于它们的总熵 $H(X_i) + H(X_j) > H(X_i, X_j)$, 因此 $T(X_i, X_j) > 0$, 表示两者之间存在信息交流或信息传递, $T(X_i, X_j)$ 越大表示观测孔 X_i, X_j 之间的联系愈密切; (3) 信息具有对称性和非负性, $T(X_i, X_j) = T(X_j, X_i) \geq 0$ 。

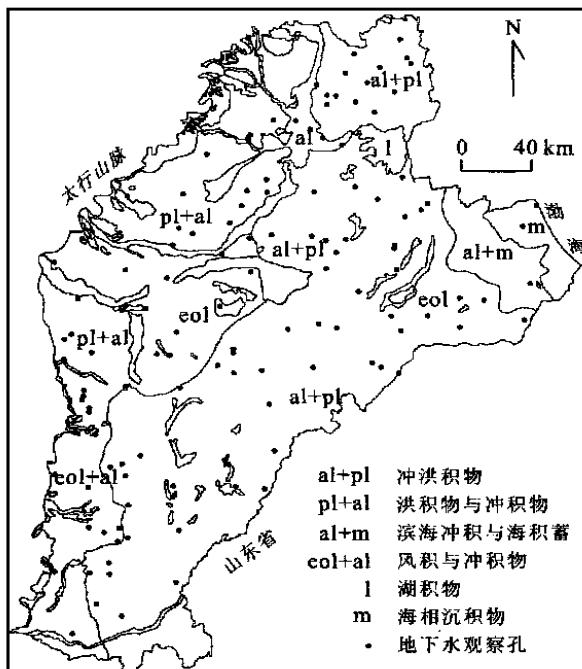


图1 河北平原地质、水文地质背景与观测孔分布

Fig. 1 Geological and hydrological background and distribution of monitoring wells in Hebei plain

3 研究实例

3.1 研究区地下水观测网背景

河北平原第四系孔隙地下水是一个具有多层次流动系统结构的区域含水系统, 分布面积超过 $6 \times 10^4 \text{ km}^2$, 区域地下水从南向北径流长度约 350 ~ 380 km, 东西向从山区补给到渤海排泄的空间距离为 200 km 左右。整个系统范围内分布地下水水位观测孔 260 个左右, 其中属于区域地下水观测孔为 140 个。根据对部分水位数据统计检验的结果, (χ^2 统计检验, 置信水平 = 0.05), 就整个水动力场随时间变化而言, 一般观测孔的水位数据可以较好地用正态分布函数来描述。图 1 表现了研究区的地质背景概况、区域地下水流向和区域地下水观测孔的空间分布, 可以直观地看出观测孔在空间上的分布并不合理, 部分地段观测孔的分布过于稀疏, 而部分地段观测孔的数量似乎又过于密集, 完全可能存在冗余性观测孔。

河北省地质局第一水文地质工程地质大队, 河北省水文总站。河北省地下水资源评价报告。石家庄, 1990。

河北省环境地质勘察院。河北平原水资源开发引起的环境地质问题及其对策研究数据库。石家庄, 1998. 12。

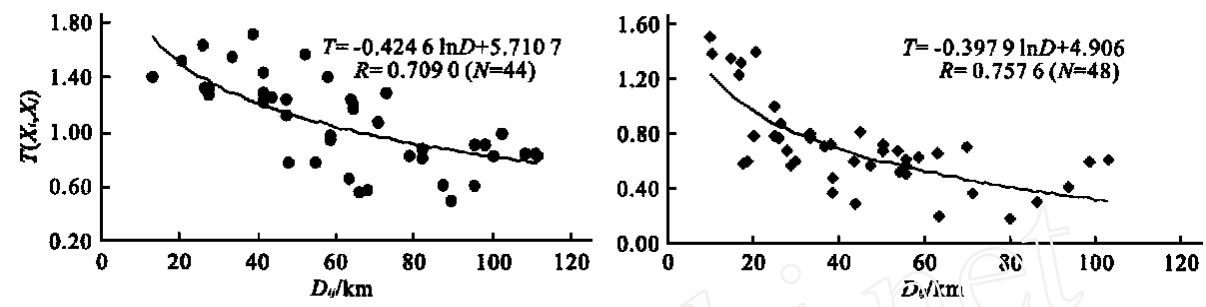


图 2 不同介质场水位信号衰减与距离的统计关系

Fig. 2 Statistical relationships between information transfer and distances in different geological deposits

表 1 不同介质场水位信号传递的变化特征

Table 1 Characteristics of information transfer between monitoring points in different geological deposits

介质场类型	冲洪积物		冲湖积物	
渗透性比较	大		小	
非均质各向异性	强		弱	
信号传递衰减规律	对数曲线		对数曲线	
信号衰减系数	0.424 6		0.397 9	
信号平均传递量	1.084 6		0.720 4	
	(44 个数据平均)		(48 个数据平均)	
信号传递 平均范围比较	最大值	1.712 7	最大值	1.504 2
	20 km	1.4 ⁺	20 km	1.0 ⁺
	40 km	1.2 ⁺	40 km	0.9 ⁺
	60 km	>1.0	60 km	<0.6
统计关系离散性	大		小	

3.2 观测孔之间信号传递距离

观测网的冗余性评价首先要确定一个观测网适宜的密度,即某个观测孔收集的数据信息能够代表的范围大小,这与观测网的精度和目的有关,同时也与观测对象变化的复杂程度有关,这是观测网设计的难题。但如果从信息传递的角度来理解,某个观测孔的信号传递距离越大,其控制的范围越广,观测孔的密度就可以减小;反之则要求加大观测孔的密度以增加新的信息来源。因此,通过评价观测孔的信号传递能力,可以很好地确定观测孔的控制范围。

观测孔的信号传递能力受多种因素影响,根据信号通讯理论,信号在传输过程中随时间和距离将产生信号衰减,信号的衰减幅度大小与其传输的距离、时间以及传输的媒质特性有关。对水位信号而言,其信号衰减随观测孔之间的空间距离有关,同时还受到含水层介质特性的影响。为了寻求观测孔水位信号衰减的规律性特征,在计算不同观测孔对之间信息传递量 $T(X_i, X_j)$ 的基础上 ($i = 1, 2, \dots, 140; j = 1, 2, \dots, 140$),研究了不同背景条件信息传递量与空间距离 $D_{i,j}$ 之间的统计关系,目的就是以

它们的统计规律性来确定观测网的适宜密度。图 2 和表 1 分别给出了位于冲洪积物和冲湖积物地区,沿地下水流向观测孔之间的水位信号传递与距离的统计关系曲线,同时也给出了曲线的拟合形式和评价结果。分析和比较这两条衰减曲线的特征,我们可以得到的基本结论是:(1)在河北平原区域地下水流动系统中,随着观测孔之间空间距离的加大,水位信号衰减基本规律是对数曲线形式;(2)含水介质的差异,包括非均质各向异性和渗透性大小都对水位信号的传递产生明显的影响,但这种影响没有改变水位信号传递的基本规律。在渗透性强的地带或区域,水位信号传递的空间范围要大于渗透性弱的地带或区域,而非均质各向异性的影响主要造成水位信号传递与空间距离的关系变化波动和随机性增大,这在图 2 已经反映出来了。以上获取的对数曲线衰减统计方程成为地下水观测孔优化的基础。

3.3 观测网冗余性观测孔的确定

明确了观测孔水位信号传递衰减的一般统计规律,意味着可以确定观测孔的一般控制范围大小。不同的信息传递标准代表不同的控制范围,具体标准值的确定应该与观测网的目的、要求和研究的尺度等不同方面联系起来,对应于精度要求比较高,小尺度研究的情况,可以提高标准以保证具有足够的监测精度,当然观测网的密度也将随之增大;对应大尺度的区域监测,标准值可以适当减少,单个观测孔的控制范围加大而观测网密度降低。因此,信息传递量标准具有一种调节器的作用和功能,不同目的和研究尺度与观测网的关系可以通过信息传递量这一指标值来进行调整和匹配,进而可以确定适宜的观测网密度。

一旦确定具体的信息传递量标准后,如何根据这一标准来计算某一具体观测孔的控制范围?前面的研究和分析结果揭示了水位信号衰减的统计规

表 2 冗余性观测孔数据
Table 2 Reduction of redundant monitoring points

观测孔对编号	X 坐标	Y 坐标	D_{ij}/m	$T_{ij}/\text{奈比}$	观测孔对编号	X 坐标	Y 坐标	D_{ij}/m	$T_{ij}/\text{奈比}$
8 - 9	2028850	4217300	25 658	1.628	11 - 172	20281050	4183000	14 847	1.026
51 - 50	20486400	4274100	5 233	1.184	198 - 199	20384500	4282425	11 702	0.999
142 - 105	20372425	4120450	15 649	1.446	215 - 208	20430725	4334625	13 438	1.305
142 - 106	20372475	4119700	14 919	1.456	215 - 210	20427825	4314350	32 376	1.101
170 - 107	20292125	4159250	2 976	1.443	215 - 216	20430350	4339750	8 945	1.264
170 - 108	20291775	4159125	3 164	1.777	211 - 213	20420425	4319300	8 927	1.056
170 - 109	20295000	4151950	10 546	1.002	221 - 223	20346075	4259100	8 475	1.438
120 - 115	20294950	4105000	17 147	1.070	289 - 229	20398100	4255930	11 986	1.058
164 - 162	20314350	4211480	3 612	1.005	227 - 231	20381300	4273300	12 039	1.248
164 - 163	20316150	4214570	36	2.036	236 - 235	20471500	4373100	0	1.911
164 - 165	20316180	4214600	10	1.467	256 - 251	20445600	4364100	16 465	1.404
164 - 150	20326800	4212100	10 920	1.237	285 - 284	20421050	4255300	8 522	0.998
164 - 152	20325850	4198780	18 547	1.032	298 - 296	20430100	4236210	0	1.251
164 - 153	20325850	4198780	18 547	1.008	301 - 302	20469650	4172560	10 475	1.378
124 - 147	20344200	4198490	16 548	1.010	301 - 303	20454700	4179350	5 950	1.247
8 - 160	20316800	4235150	41 125	1.445	336 - 334	20377000	4185670	12 236	1.021

观测孔对编号 8 - 9 表示观测孔 No. 8 与 No. 9 ,后者属于冗余观测孔 ;坐标为冗余观测孔的位置.

律,同时也发现由于第四系介质场的非均质各向异性对信号衰减的影响,水位信号虽然遵循对数曲线衰减的基本规律,但在不同的空间位置,不同的流动方向,观测孔之间的信号传递和衰减存在着比较大的差异.这种差异的存在导致我们不便将整个流动系统视为一个集中参数系统,基于所有观测孔的统计信息联系去建立一个适应整个区域流动系统的统计方程,这种方式处理的结果将会模糊由于介质场的变化而存在衰减系数的差异,尽管所建立的统计方程对系统整体而言具有统计意义,但对每一个局部将可能出现一定的误差.因此,在具体利用水位信号衰减的基本规律进行观测网结构的信息评价和扩充时,应分别考虑冲洪积扇和冲湖积物分布区观察孔监测能力的差异性.

本次研究对象为整个河北平原地下水系统,研究尺度属于区域性,因此,在比较了不同信息传递标准与观测网的冗余成分后,选择观测孔之间的信息传递量 1.0.如此,在冲洪积扇分布区观测孔的空间控制范围 < 60 km,而湖积物地段 < 20 km,这个空间范围即是以某一观测孔为中心的冗余信息搜索域.对该范围所有观测孔之间的信息传递关系进行评价,评价内容包括(1)计算每一对观测孔之间的空间距离和(2)相互之间的信息传递量大小.在该搜索域内,如果存在另一或多个观测孔并且它们之间的信息传递量大于给定标准的话,表明观测孔之间具有良好的信息传递,这些观测孔所提供的水位数据

信息可以彼此相互取代,除保留其中某个观测孔外,其他均属于冗余性的观测站点都可以删除.保留观测孔的选择原则是:(1)在整个信息搜索域内的信息熵最大;(2)适宜的空间位置和其他特殊要求限定.如果观测孔之间信息传递量小于给定的标准值,这些观测孔均应保留,因为它们均是地下水变化的信息来源.这种方法的优点可以体现几个方面:用信息传递量这一指标大小来评价观测孔之间的相互关系,可以发现所有那些属于冗余性的观测站点,使优化结果具体化;同时信息传递标准值为一客观的控制性指标,并没有“平均化、等密度”的方法处理非均质各向异性问题,不同背景、不同地段可以出现不同的观测孔密度,保证在观测孔数量优化的同时不会增加新的其他不确定成分;此外这种方法也允许研究者根据某些条件和需求保留特定的观测孔.

根据上述计算过程对河北平原区域流动系统现有观测站点信息冗余性评价(因数据量太大,不便列出),结果表明目前观测网的冗余性问题还是相当严重的.以信息传递量 $T(X_i, X_j) = 1.0$ 作为评价标准,所有的 140 个观测孔中有 36 个观测孔的数据基本没有提供新的信息,属于冗余性观测孔,表 2 列出了这些冗余性观测孔的基本数据.可以看出,在 36 对数据中有 4 对观测孔的空间位置基本重叠,而确定的冗余观测孔最远距离可达 41 km(No. 8 - No. 160),实际情况是这一对观测孔均分布在洪积扇顶部透水性强的地段.在比例上这些冗余观测孔占全部观测

表 3 优化前后地下水不同水位高程的分布面积统计对比

Table 3 Comparison of different water levels areas in prior and post of removed redundant monitoring points

水位区间/m	- 40 - 30	- 30 - 20	- 20 - 10	- 100	010	1020	2030	3040	4050	5060	6070	7080	合计
原面积/ km ²	1 952	4 616	7 998	12 846	9 558	7 848	7 130	5 180	2 300	1 400	704	506	62 038
现面积/ km ²	1 968	4 386	7 675	13 216	9 674	7 416	7 581	5 148	2 374	1 345	708	547	62 038
误差/ %	0.81	5.24	4.21	2.80	1.20	5.83	5.95	0.62	3.12	4.09	0.56	7.50	3.46

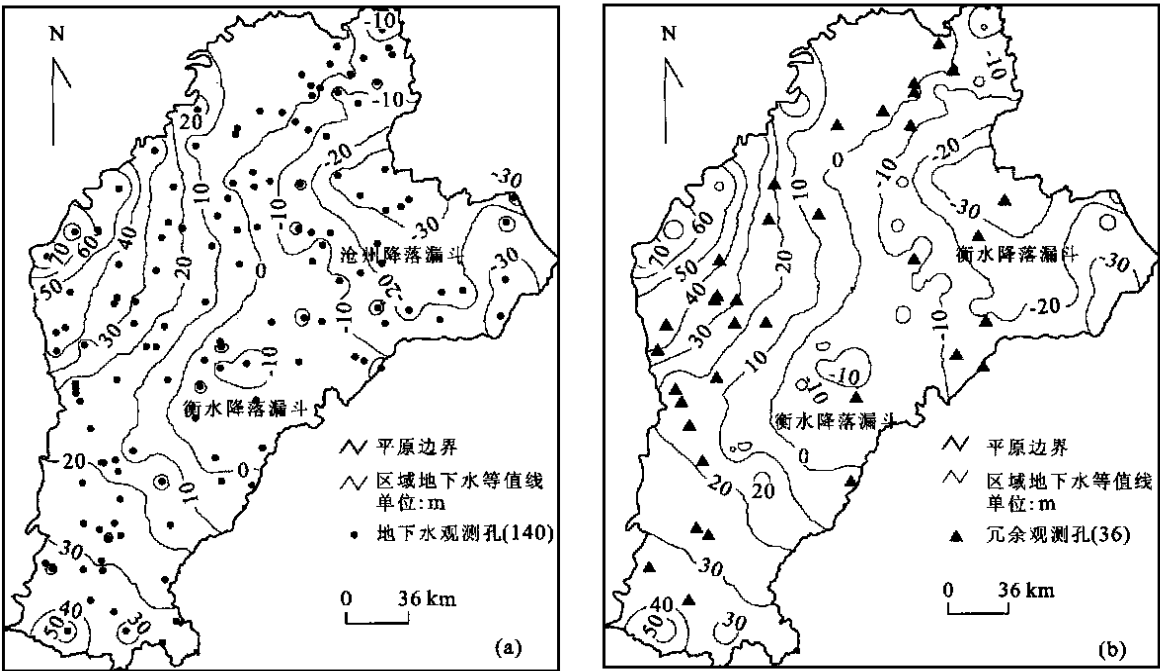


图 3 优化前后地下水水位高程

Fig. 3 Groundwater levels before (140 points) (a) and after (104 points) (b) removing redundancy

a. 优化前, 140 个观测孔; b. 优化后, 104 个观测孔

孔数量的 $36/140 = 25.71\%$, 这意味着删除这些观测孔, 仅在经济方面可以节省 26 % 的经费支出。

为了证实这些冗余观测孔的“作用”, 验证其提供的水位数据是否包含有不能忽略的信息成分, 利用 GIS 工具进行水位空间形态方面的比较和分布面积变化的统计对比。图 3a 和 3b 分别表现了基于不同观测孔数量 (140 个和 104 个) 空间插值的地下水水位形态, 可以看出两者空间分布非常相似, 几乎没有实质性的区别, 而且后者还消除了部分异常变化成分; 同样借助 GIS 的空间分析工具, 具体统计了各自不同水位高程的分布面积 (表 3)。因此, 根据信息熵方法所确定的 36 个观测孔确实属于冗余性观测孔, 因为它们的去除基本没有影响河北平原区域地下水水位空间分布的形态。

4 结论

(1) 基于信息熵的概念和方法来解决观测网系统的数据冗余问题是一种值得深入研究的技术方法, 这种方法突破了传统的思维方法, 不是仅从“数据、精度”等方面的统计评价, 而是从观测网的系统性和彼此信息关联的角度评价观测网提供信息的能力, 从观测孔的信息熵和它们之间的信息关系, 可以具体确定每一观测孔提供信息的能力, 据此来解决观测网存在的信息冗余问题; (2) 冗余性观测孔的去除并不意味着增加了观测网的精度, 冗余性的消除主要是减少数据噪音和节省了观测网的经济投入, 这是在保证原有观测网数据信息基本不损失的前提下实现的。需要强调的是, 除数据冗余外, 观测网可能还存在因空间结构不合理, 导致部分地段信息缺乏等问题, 因此, 真正一个优化的观测网还必须评价

观测网空间布局的合理性,还必须从结构上优化,即消除冗余性的同时也应考虑因信息不足而对观测网的扩充问题;(3)观测网的冗余性除了在观测孔数量方面表现外,同时在观测时间频率方面也可能存在冗余性,这与所监测的水位或水质变化大小有关。例如变化剧烈的对象需要加密观测次数,反之可以减少观测的频率。时间方面数据的冗余性对地下水水质观测网而言也具有重要的经济意义,目前水质分析的经济成本较高,优化监测的时间频率可以使观测网的投入与产出比更高,这方面同样可以借用信息熵理论和方法来实现;(4)信息传递量标准可以作为一种调节器,不同目的和空间尺度的观测网对应于不同的观测要求,这种关系可通过信息传递量这一标准进行匹配,以确定适宜的观测网密度。但信息传递量标准的确定需要综合多种因素来考虑,它涉及到观测网目的、要求和研究尺度,涉及到对象背景条件的复杂程度等许多因素,需要以后结合不同条件和要求来进一步深入研究。

参考文献:

- [1] Morin G, Fortin J P, Sochanska W, et al. Use of principal component analysis to identify homogeneous precipitation stations for optimal interpolation [J]. *Water Resources Research*, 1979, 15(6): 1841 - 1850.
- [2] Galeati G, Rossi G, Pini G, et al. Optimization of a snow network by multivariate statistical analysis [J]. *Hydrol Sci J*, 1986, 31(1): 93 - 108.
- [3] Burn D H, Goulter I C. An approach to the rationalization of stream flow data collection networks [J]. *J of Hydrology*, 1991, 122: 71 - 91.
- [4] Timothy B, Spruill C, Lucila C. Two approaches to design of monitoring networks ground water [J], 1990, 28(3): 430 - 442.
- [5] Haber F. An introduction to information and communication theory [M]. New York: Addison Wesley Publishing Company, 1974.
- [6] Harmancioglu N B, Necdet A. Water quality monitoring network design: a problem of multi-objective decision making [J]. *Water Resources Bulletin*, 1992, 28(1): 179 - 192.
- [7] Ozkul S. Entropy-based assessment of water quality monitoring networks [J]. *Journal of Hydrologic Engineering*, 2000, 5(1): 90 - 100.

ENTROPY-BASED APPROACH TO REMOVE REDUNDANT MONITORING WELLS IN REGIONAL-SCALE GROUNDWATER SYSTEM IN HEBEI PLAIN, CHINA

Chen Zhihua, Ding Guoping

(*Engineering Faculty, China University of Geosciences, Wuhan 430074, China*)

Abstract : Removing the redundancy is one of the purposes of optimizing monitoring network, for redundant data made the data noises and increased the operation cost. The main solution to remove redundancy is reduction of the number of monitoring wells under the precondition: there is no or few effect on the ability of information collected by network. A new approach based entropy for optimization was presented in this paper. This procedure is a three-phase method, in which the entropy was employed to measure the ability of individual station and the information transfer coefficient between well pairs was considered as a measurement of information relationships. The different statistical relationships were found between the information transfer coefficients and the distances among wells located at different geological deposits, which is the base to determine the redundant wells. This approach was demonstrated using the data from the regional-scale groundwater flow system in Hebei plain, China, 36 wells in 140 wells distributed whole area were identified as the redundant wells. And the abilities of information collected were almost the same as that before even though the 36 wells were removed.

Key words : monitoring network; redundant data; optimization; entropy; information transfer.