

文章编号: 1673 - 9469(2014) 04 - 0040 - 06

doi: 10. 3969/j. issn. 1673 - 9469. 2014. 04. 011

卧龙湖矿 8101 面地下水水质分类判别研究

姚辉 李朝阳 张永生

(合肥工业大学 资源与环境工程学院 安徽 合肥 230009)

摘要: 为分析卧龙湖 8101 工作面涌水补给源以及集水形式,统计整理了本矿井地下水样品常规水质分析数据 62 组。利用多元统计组合模型与 maltab 方法相结合对 8101 工作面涌水水源进行分析研究。结果说明: 8101 工作面涌水水源与本矿 10 煤砂岩水质类型一致。卧龙湖矿各层地下水水质的首要因子为 Ca^{2+} 、 Mg^{2+} 、 SO_4^{2-} ,体现太灰水的外来补给;第二因子是阳离子交替吸附作用,其涌水体现了地下水静储量消耗。多元统计模型与 maltab 方法相结合是一种稳定性高且切实有效,适用于地下水水源水化学分析的方法。

关键词: 地下水系统;水化学分类;主成分分析;聚类分析;贝叶斯判别分析; Matlab 软件
中图分类号: X824 文献标识码: A

Study on discrimination of Wolong Lake coal mine 8101 face of groundwater quality classification

YAO Hui, LI Chao - yang, ZHANG Yong - sheng

(School of Resources and Environmental Engineering, Hefei University of Technology, Anhui Hefei 230009, China)

Abstract: For the analysis of Wolong Lake 8101 working face water inrush source and water collection, statistics of the mine groundwater samples of conventional water quality analysis data of 62 groups were collected. Using a combination of models and multivariate statistical methods combining maltab, this paper illustrated the gushing water of 8101 Face. The results show that the water resource of 8101 consistent with the mine 10 coal sandstone type. The primary factor of Lake Wolong ore stratum of groundwater quality is Ca^{2+} 、 Mg^{2+} 、 SO_4^{2-} , which reflects Taiyuan limestone water supply being from the other water resource; the second factor is the cation exchange adsorption, which embodies the static storage consumption of groundwater flow. The method of multivariate statistical model and maltab is a kind of high stability and effective, which is applicable to discriminate the groundwater source water chemistry.

Key words: groundwater system; hydrochemical identification; principal component analysis; cluster analysis; Bayes discriminant analysis; Matlab software

地下水系统的水化学分类以及判别分析是水地质学的重要内容之一,为了快速判别突水水源,实现煤炭生产的高产高效,许多学者从不同侧面研究水化学模型,取得了一定的效果^[1-4],其中最基本方法之一是多元统计方法^[5],被证实具有一定的稳定性。但每种方法都有局限性,考虑到矿井地下水运动的空间和时间变异特性,如何提高水源识别分析精度仍然是一个需要长期解决的问题。本文以卧龙湖矿 8101 工作面为例,先采用

多元统计方法建立水化学判别模型,主要包括主成分分析、聚类分析、判别分析,建立水化学判别模型,对未知水样做出判定。同时与 Maltab 编程方法相结合^[6-8],对未知水样进行判断,二者进行对比。

一 矿区地质、水文地质条件

卧龙湖煤矿位于安徽省濉溪县铁佛、岳集境内,东距百善煤矿约 15 km,北以省界与河南省永

收稿日期: 2014 - 09 - 04

基金项目: 安徽省科技攻关项目(11010401015)

作者简介: 姚辉(1990 -),男,海南海口人,硕士,从事地质工程方面研究。

城市毗邻,南北长约 8 ~ 9 km,东西宽约 3.5 ~ 4 km,面积 28.85 km²。该煤田位于华北板块南部,东与郟庐深大断裂相距较近,属于高冀构造区东缘永夏断隆带。本区属华北地层区、鲁西分区、徐州小区。缺失上奥陶统至下石炭统和三叠系至古新统等地层,区域内除灵壁、泗县、濉溪、涡阳的石弓、龙山等地有震旦、寒武、奥陶系地层及烈山、白土有石炭、二叠系地层出露外,地表被新近系和第四系松散沉积物覆盖。地层由老至新有青白口系、震旦系、寒武系、奥陶系、石炭系、二叠系、侏罗系、白垩系、古近系、新近系及第四系。含水层从上到下有第四系、第三系冲积含水层组,煤系地层砂岩含水层组和灰岩岩溶裂隙含水层组。其中孔隙含水层组共有 4 个含水层和 3 个隔水层。据抽水试验资料,主采煤层顶底板砂岩裂隙含水层 $q = 0.004\ 012 \sim 0.009\ 37\ \text{L/s} \cdot \text{m}$ 。井下揭露的突水点变化规律,一般是开始涌水量较大,随时间延长,衰减较快,呈淋水或滴水状态。总的来说煤系砂岩裂隙含水层富水性较弱,因此,在穿过坚硬砂岩时,必须注意储存水量突然喷出。

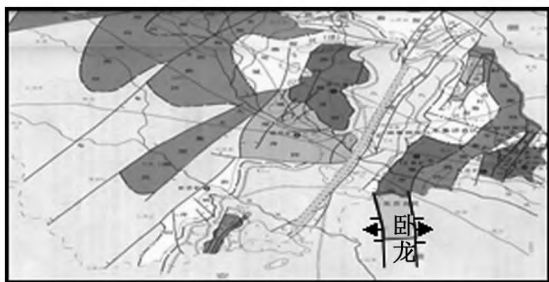


图1 矿区所在区域构造位置

Fig.1 The tectonic location of mining area

二 多元统计模型

收集了本矿井各地层地下水样品常规水质分析数据 62 组,水样采集于水源井、抽水孔及采掘

工作面,时间主要是 2012 - 2013 年,此期间矿区总体开采规模不大,地下各含水层未出现较大水位变化,故本次研究从静态角度对水质资料进行多元统计分析。

2.1 地下水化学的主成分分析

地下水重分布最广,含量较高的 7 种离子: Na^+ 、 K^+ 、 Ca^{2+} 、 Mg^{2+} 、 Cl^- 、 SO_4^{2-} 、 CO_3^{2-} ,它们的不同浓度,组成反映了不同的地下水水质特征,其浓度差异是地下水化学分类的主要依据。这些离子并非完全独立地存在于地下水体系中,它们之间总有些相关性,离子之间的相关意味着其反映的地址信息重合。若不作处理,就进行聚类分析,相当于给这些离子变量进行了加权,夸大了某些地址信息因素的作用,使计算结果发生畸变。为克服这种现象,本文先对离子进行主成分分析,找出基本的数据结构,用新生成的主成分得分作为新变量进行聚类分析。利用 spss 软件的 Na^+ 、 K^+ 、 Ca^{2+} 、 Mg^{2+} 、 Cl^- 、 SO_4^{2-} 和 CO_3^{2-} 的相关系数矩阵(表 1)。由表 1 得出: Ca^{2+} 和 Mg^{2+} 、 Na^+ 、 K^+ 与 Cl^- 直接存在较强的相关关系,其他离子变量直接也存在一定相关关系,需要对原始变量数据进行因子分析。

利用因子分析中的主成分分析法,从离子相关系数矩阵中提取初始因子,得到了相关系数矩阵的 6 个特征值(主成分所解释的方差)。按照提取主成分的个数一般要求其累计方差贡献超 80% 的原则,提取前 3 个主成分,其累计方差贡献率达 89.661% (见表 2)。

选取的因子初始载荷矩阵一般需要进行方差极大正交旋转,使各离子变量在同一因子上的载荷区别明显。表 3 为采用方差极大正交学专方法迭代 3 次后的因子载荷矩阵,每个离子变量的信息被保留的程度体现在各自的公因子方差上。

表 1 相关系数矩阵

Tab.1 Correlation matrix

	Na^+	Ca^{2+}	Mg^{2+}	Cl^-	SO_4^{2-}	CO_3^{2-}
$\text{Na}^+ + \text{K}^+$	1.000	-.424	-.505	-.037	.331	.433
Ca^{2+}	-.424	1.000	.877	.566	.443	-.420
Mg^{2+}	-.505	.877	1.000	.488	.351	-.403
Cl^-	-.037	.566	.488	1.000	.314	-.208
SO_4^{2-}	.331	.443	.351	.314	1.000	-.552
CO_3^{2-}	.433	-.420	-.403	-.208	-.552	1.000

表2 解释的总方差

Tab.2 Total variance analysis

成份	初始特征值			提取平方和载入			旋转平方和载入		
	累积 %	合计	方差 %	累积 %	合计	方差 %	累积 %	合计	方差 %
1	3.04	50.71	50.71	3.04	50.71	50.71	2.19	36.47	36.47
2	1.40	23.23	73.94	1.40	23.23	73.94	1.63	27.11	63.58
3	0.94	15.73	89.66	0.94	15.73	89.66	1.57	26.08	89.66
4	0.48	7.99	97.65						
5	0.12	1.91	99.55						
6	0.03	0.45	100.00						

表3 旋转因子载荷矩阵

Tab.3 Rotated component matrix

类别	成份			公因子方差
	1	2	3	
Na ⁺	-0.152	-0.040	0.978	0.980
Ca ²⁺	0.829	0.323	-0.311	0.889
Mg ²⁺	0.780	0.275	-0.422	0.862
Cl ⁻	0.862	0.037	0.147	0.767
SO ₄ ²⁻	0.337	0.824	0.420	0.969
CO ₃ ²⁻	-0.107	-0.874	0.369	0.912

表4 主成分得分

Tab.4 The score of main component score

编号	f1	f2	f3	编号	f1	f2	f3	编号	f1	f2	f3
1	-0.172	-0.344	-1.716	22	-0.142	0.488	1.443	43	-0.258	-0.143	-1.345
2	0.856	-1.033	-0.465	23	-1.510	0.270	-1.521	44	-0.201	0.906	0.208
3	0.386	-0.469	-1.181	24	-0.416	-1.084	-0.482	45	-0.798	1.151	0.672
4	-0.484	0.849	-0.808	25	0.379	-2.268	1.394	46	-0.821	1.246	0.548
5	0.255	-0.337	-0.980	26	-1.035	1.381	0.433	47	0.292	0.488	0.985
6	-0.433	-0.002	-1.505	27	-0.535	0.834	0.245	48	-0.375	0.747	0.569
7	-0.319	0.003	2.209	28	0.664	-2.193	0.842	49	-0.781	-0.099	-1.395
8	0.291	-1.596	1.429	29	-0.394	-0.763	-0.405	50	-0.458	0.775	0.235
9	-0.359	-1.296	-0.067	30	0.437	-2.292	0.670	51	-0.563	0.907	0.438
10	0.035	0.237	-1.768	31	-0.626	0.191	-0.037	52	2.680	0.870	-1.067
11	-0.106	0.280	-1.756	32	-0.453	0.182	0.056	53	-1.334	-0.916	-0.928
12	-0.865	-0.595	-1.135	33	-0.440	0.265	0.146	54	0.498	-0.643	0.352
13	-0.247	-0.155	1.649	34	0.584	-0.284	-0.885	55	1.455	-1.604	0.720
14	-0.807	-0.462	-0.435	35	-0.477	0.073	0.030	56	2.614	0.792	0.472
15	-0.563	-0.379	-0.127	36	-0.554	-0.255	-0.865	57	1.246	0.140	1.669
16	-0.562	-0.091	-0.171	37	-0.630	0.832	0.372	58	-0.011	1.058	-0.988
17	0.511	-1.954	1.249	38	-0.527	0.823	0.464	59	0.344	-0.681	0.720
18	-0.767	-0.013	-0.183	39	-0.575	1.098	0.827	60	3.413	0.616	-1.061
19	-0.645	-0.523	-0.684	40	-0.644	2.328	2.655	61	2.551	1.675	-0.053
20	-0.816	0.453	0.365	41	-0.247	0.349	0.323	62	2.550	1.513	-0.768
21	0.138	-1.656	0.014	42	-0.226	0.310	0.381				

2.2 水质样品的系统聚类分析

聚类分析是研究事物分类的一种方法,将一批样本或变量按照他们在性质上的亲疏程度加以分类。为获取矿区地下水各子系统典型水样,利用系统聚类法中的最短距离法、最长距离法、类平

均法、重心法和离差平方和法等5种方法对水样进行聚类,距离测度使用平方欧式距离,聚类变量即为正交因子(TDS因子和碱度因子)。在上述主成分分析的基础上,利用旋转后的因子回归计量值,即主成分1、主成分2和主成分3的得分记为 f_1 、 f_2 、 f_3 ,代替原始水样数据用于聚类分析进行系

统聚类分析。用欧式距离表示主成分得分值之间的相似性,即

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}$$

其中 i, k 均为样品序号 $i = 1, 2, 3, \dots, 62; j$ 为样品指标数 $j = 1, 2, 3$ 。主成分得分值见表 4。

利用 spss 软件对 62 个水样的主成分得分进行了聚类分析,聚类结果如图 2 所示。若将并类距离 d 设为 5,则可以将 62 组水样分为 4 类,即 1, 2, 3 为一类,4, 5, 6 分别自成一类。如图两个椭圆区域所示,这两个椭圆区域中 8101 工作面的水样为本矿 10 煤和太灰。根据这聚类结果可判定,卧龙湖 8101 面部分出水与本矿 10 煤和太灰水质相似性高,地下水水质类型较为相似。

2.3 地下水的 Bayes 判别分析

利用系统聚类得到的地下水各含水层的典型水样,先剔除水样类别错误的,选择剩下的 44 组水样,建立判别函数。Bayes 线性判别一般要求母体变量观测值必须服从多元正态分布,而且要求各母体的协方差阵无显著差别。在应用中,当各母体的样本规模比较接近时,如最大一组的样本容量不超过最小一组样本容量的 1.5 倍时,违反协方差阵相等的假设条件,影响也不太大。表 5 为函数系数。

其中 y_6, y_7 分别表示矿区第四系和煤系水的判别函数。判断结果与聚类结果一致,接近 100% 的准确率,说明建立的判别函数能很好的预测突水水样的类别。判别函数代入 $Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, CO_3^{2-}$ 的浓度即可。

3 Matlab 软件在多元统计分析的应用

在 Matlab 较早的版本中,统计功能还不是那么强大,而在 Matlab 6. x 版本中,仅在统计工具箱 (Statistic Toolbox) 中的功能函数就达到 200 多个,

功能已足够赶超任何其他专用的统计软件,Matlab 统计工具箱几乎包括了数理统计方面的书有概念、理论、方法、算法以及实现。在应用上,Matlab 具有其他软件不可比拟的操作简单,接口方便,扩充能力强等优势。因此可以预见该软件在统计应用上的重要地位,利用卧龙湖矿及周边矿的水样以实例给出 Matlab 在主成分分析、聚类分析、判别分析的应用。

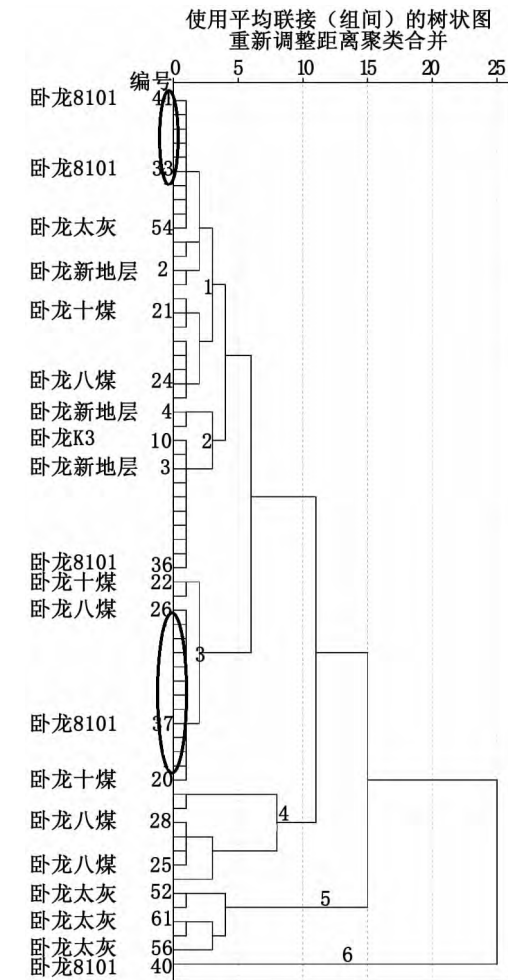


图2 聚类分析
Fig.2 Cluster analysis

表 5 函数系数

Tab.5 The function coefficient

	组数						
	1	2	3	4	5	6	7
Na ⁺	-0.023	1.605	-0.064	0.121	0.059	0.063	-0.011
Ca ⁺	-1.163	1.968	-0.962	0.300	0.355	-0.207	0.126
Mg ²⁺	2.396	0.744	2.057	-0.254	-0.485	-0.035	0.429
Cl ⁻	2.112	-0.126	2.079	1.614	2.132	2.422	3.316
SO ₄ ²⁻	0.189	-0.991	0.161	0.215	0.280	0.302	0.329
CO ₃ ²⁻	0.520	-0.301	0.359	0.674	0.862	0.433	0.656
(常量)	-13.200	-24.680	-10.224	-12.296	-16.273	-12.131	-20.291

3.1 主成分分析

利用 Matlab6.5 中的 princomp 命令实现,具体程序如下:

```
x = [ 9.432 20.265 13.177 17.836 14.179
11.015 45.792 43.938 24.186 8.055 8.251
14.362 40.611 21.273 23.434 23.179 35.058
23.147 18.077 28.821 23.625 36.627 12.464
19.793 37.334 28.993 ... ..14.44 27.55 10.19
9.62 9.16 11.79 10.27 8.85 6.92 6.84 6.62 6.91
8.98 9.10 9.06 6.76 7.22 5.99 7.29 7.07 6.55
5.70 6.31 2.66 20.57 12.06 17.32 4.72 10.15
3.05 15.48 1.66 2.75 2.51 ]
```

```
x = x';
```

```
stdr = std(x); % 求各变量标准差
```

```
[n m] = size(x);
```

```
sddata = x./stdr(ones(n,1)); % 标准化
```

变换

```
[p princ egnvalue] = princomp(sddata) % 调用主成分分析程序
```

```
p3 = p(:,1:3) % 输出前三个主成分系数
```

```
sc = princ(:,1:3) % 输出前三个主成分得分
egvalue % 输出特征根
```

```
per = 100* egvalue/sum(egvalue) % 输出各个主成分贡献率
```

执行后得到所要的结果,这里事前三个主成分、主成分得分、特征根,即

$$p3 = \begin{bmatrix} 0.2735 & 0.7204 & -0.1766 \\ -0.5266 & -0.0406 & -0.2130 \\ -0.5121 & -0.1467 & -0.1901 \\ -0.3620 & 0.2170 & -0.5660 \\ -0.3194 & 0.6403 & 0.3045 \\ 0.3905 & -0.0269 & -0.6887 \end{bmatrix}$$

$$\text{egvalue} = \begin{bmatrix} 3.0421 \\ 1.3939 \\ 0.9436 \\ 0.4792 \\ 0.1145 \\ 0.0268 \end{bmatrix}$$

```
per = [50.7009 23.2316 15.7266 7.9866
1.9076 0.4468]
```

```
Z1 = 0.2735 × 1 - 0.5266 × 2 - 0.5121 × 3 -
0.3620 × 4 - 0.3194 × 5 + 0.3905 × 6;
```

```
Z2 = 0.7204 × 1 - 0.0406 × 2 - 0.1467 × 3
```

```
+ 0.2170 × 4 + 0.6403 × 5 - 0.0268 × 6;
```

```
Z3 = -0.1766 × 1 - 0.2130 × 2 - 0.1901 × 3
- 0.5660 × 4 + 0.3045 × 5 - 0.6887 × 6;
```

第一主成分贡献率为 50.70%, 第二主成分贡献率 23.23%。第三主成分贡献率为 15.73%。前三个主成分累计贡献率达 89.66%。与 spss 软件运算得到方差贡献率(89.661%)基本一致。

3.2 聚类分析

利用 Matlab6.5 中的 cluster 命令实现,具体程序如下

```
x = [ 9.432 2.28 4.56 3.38 3.51 10.02 ... ..
```

```
11.015 1.60 3.27 3.46 5.60 7.77
```

```
45.792 0.16 0.12 2.79 25.37 21.84
```

```
43.938 0.61 0.30 4.62 0.67 21.28
```

```
7.398 19.85 7.72 9.35 24.49 1.66
```

```
20.750 14.59 10.39 6.71 37.97 2.75
```

```
14.029 17.79 9.84 6.28 33.87 2.51 ];
```

```
[n m] = size(x);
```

```
stdr = std(x);
```

```
xx = x./stdr(ones(n,1)); % 标准化变换
```

```
y = pdist(xx); % 计算个样本间距离(这里为欧氏距离)
```

```
z = linkage(y); % 进行聚类(这里为最短距离法)
```

```
t = cluster(z,4) % 将全部样本分为4类
```

```
find(t==2); % 找出属于第2类的样品编号
```

```
h = dendrogram(z); % 画聚类谱系图
```

执行后得到所要结果,聚类谱系图见图 3。

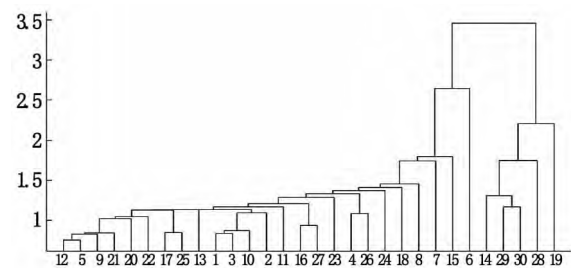


图3 聚类谱系图

Fig.3 Cluster analysis

4 结论

1) 在保证原始数据信息损失最小的前提下,通过主成分分析法,前 3 个主成分(50.70%, 23.23%, 15.73%),其累计方差贡献率达 89.661%。

2) 聚类分析方法, 得出 8101 工作涌水水源与本矿 10 煤砂岩水质类型一致。

3) 因子分析方法, 得出卧龙湖矿各层地下水水质的首要因子为 Ca^{2+} 、 Mg^{2+} 、 SO_4^{2-} , 体现太灰水的外来补给; 第二因子是阳离子交替吸附作用, 其涌水体现了地下水静储量消耗。

4) 建立的 Bayes 线性判别模型, 具有计算简便、误判率低、稳定性高等特点。同时运用 Matlab 程序, 解释的结果与多元统计模型研究结果总体上一致, 且与该矿区水文地质条件相吻合。

模型的建立依赖于水样的数量、变量的选择、总溶解固体物、硬度、碱度等, 建议今后采集更多的地下水样品, 从中选择时间跨度较小且空间分布均匀的水样作为样本, 模型需不断调整和完善。

参考文献:

[1] 尹国勋, 杨娜, 贺玉晓, 等. 焦作市市区地下水水质现

状评价[J]. 环境工程, 2004, 22(4): 66 - 69.

[2] 贲旭东, 郭黄海, 解奕伟, 等. 模糊综合评判在矿井突水水源判别中的应用及探讨[J]. 矿业安全与环保, 2006, 33(3): 57 - 59.

[3] 王广才, 王秀辉, 李竞生, 等. 平顶山矿区矿井突(涌)水水源判别模式[J]. 煤田地质与勘探, 1998, 26(3): 47 - 50.

[4] 闫志刚, 杜培军, 郭达志. 矿井涌水水源分析的支持向量机模型[J]. 煤炭学报, 2007, 32(8): 842 - 847.

[5] 陈桂明. MATLAB 数理统计(6. x) [M]. 北京: 科学出版社, 2002.

[6] 刘则毅. 科学计算技术与 Matlab [M]. 北京: 科学出版社, 2001.

[7] 李涛. Matlab 工具箱应用指南—应用数学篇 [M]. 北京: 电子工业出版社, 2000.

(责任编辑 王利君)

(上接第 39 页)

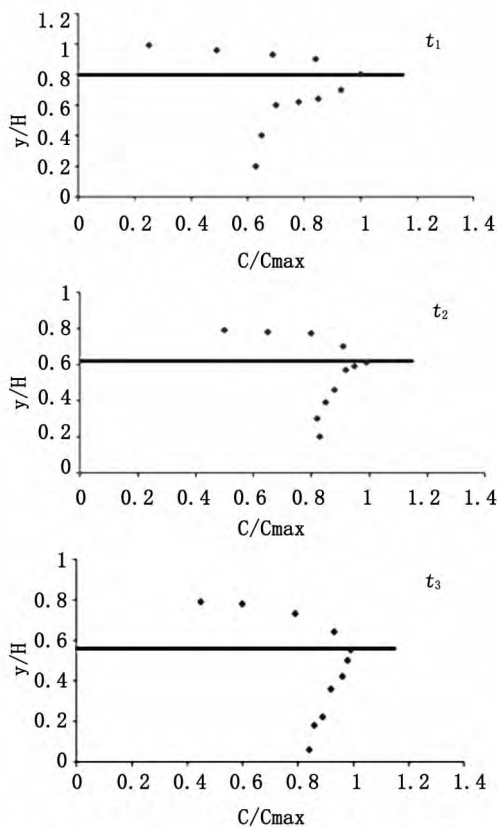


图4 随时间的增加污染物浓度垂向分布图

Fig.4 The vertical distribution of pollutant concentration as time goes on

3 结论

在植物的作用下, 水流的紊动特性增强, 增加了污染物的垂向扩散作用。而且, 水流流速减慢, 污染物的浓度在植物的作用下能更快的趋于均匀, 缩短了污染物的垂向混合时间, 截留了部分污染物质, 使下游水体的污染物浓度降低, 从而有效控制了下游水体的污染程度。

参考文献:

[1] NEPF H M, SULLIVAN J A, ZAVISTOSKIV R A. A model for diffusion within emergent vegetation [J]. Limnol Oceanogr, 1997, 42(8): 1735 - 1745.

[2] NEPF H M, MUGNIER C G, ZAVISTOSKI R A. The effects of vegetation on longitudinal dispersion [J]. Estuarine, Coastal and Shelf Science, 1997, 44: 675 - 684.

[3] TANINO Y, NEPF H M. Lateral dispersion in random cylinder arrays at high reynolds number [J]. Journal of Fluid Mechanics, 2008, 600: 339 - 371.

[4] TANINO Y, NEPF H M. Laboratory investigation of lateral dispersion within dense arrays of randomly distributed cylinders at transitional reynolds number [J]. Physics of Fluids. 2009, 21(4): doi: 10.1063/1.161.3119862.

[5] 姜新佩, 张欢, 李莹. 人工湿地中植物对生活污水的净化研究[J]. 河北工程大学学报: 自然科学版, 2014, 31(2): 59 - 63.

(责任编辑 刘存英)