

# 基于卡尔曼滤波和模糊集的地下水污染羽识别

江思珉, 张亚力, 周念清, 赵 珊

(同济大学 土木工程学院, 上海 200092)

**摘要:** 在场地信息不确定性的前提下基于卡尔曼滤波技术和模糊集合理论提出一种地下水污染羽识别的方法. 利用卡尔曼滤波技术通过连续作用采样点更新综合污染羽与误差协方差矩阵, 并结合误差协方差矩阵和不确定度的关系进行新采样点的选取; 引入模糊集理论进行污染羽的模糊对比, 以确定污染源的权重. 算例研究表明, 该方法是一种有效的污染羽识别算法, 能够以尽可能少的采样点确定污染源的位置, 并较为准确地达到识别污染羽的目的.

**关键词:** 地下水污染; 污染羽识别; 卡尔曼滤波; 模糊集

**中图分类号:** P641.2, P641.8

**文献标志码:** A

## Groundwater Contaminant Plume Identification by Using Kalman Filter Technique and Fuzzy Set Theory

JIANG Simin, ZHANG Yali, ZHOU Nianqing, ZHAO Shan

(College of Civil Engineering, Tongji University, Shanghai 200092, China)

**Abstract:** A methodology is proposed to identify the contaminant plume based on Kalman filter technique and fuzzy set theory. In this methodology, the Kalman filter technique is adopted to update the composite contaminant plume and the corresponding error covariance matrix by using the sampling points sequentially, and the relationship between error covariance matrix and uncertainty of the plume is then combined to select a new sampling point; the fuzzy set theory in the methodology is introduced to update the weight of the potential source location through the comparison of the updated composite plume and the individual plume. The case study indicates that the proposed methodology is effective to identify the contaminant plume, and the number of sampling points is close to a minimum value.

**Key words:** groundwater contamination; contaminant plume identification; Kalman filter; fuzzy set

我国地下水污染状况已十分严重. 与地表水体相比, 地下水一旦遭受污染其治理和修复通常需要更长的时间(常持续 30 年以上), 其治理和修复费用往往极为昂贵. 在地下水污染治理和修复过程中往往需要有污染羽的时空描述资料. 对污染羽进行时空描述通常需要通过建立相应的地下水污染监测网, 以获取地下水的物理特性、化学特性和生物特性在时间、空间上的变化特征数据<sup>[1]</sup>.

在地下水水质的监测过程中, 过量取样和取样不足是比较常见的问题. 过度取样会造成不必要的取样和分析费用, 而取样不足则不能获得污染物随时空变化的准确信息, 进而不能确立地下水污染治理和修复的合理方案<sup>[2]</sup>. 因此, 文献<sup>[3-7]</sup>通过对地下水污染监测网进行优化设计, 可以在不降低监测精度和不影响评价风险的前提下减少不必要的取样或增加必要的取样, 从而以尽可能少的费用准确地识别出地下水污染羽.

地下水系统是一个受多种因素影响的复杂系统, 存在着各种不确定性. 诸多不确定性的存在使得很难准确地刻画地下水中污染物的空间分布特征, 主要的不确定性因素包括场地信息的不确定性和污染源(位置和强度)的不确定性等<sup>[8-10]</sup>. 在污染源位置未知和场地信息不确定性的前提下, 本文基于卡尔曼滤波技术和模糊集合理论提出一类地下水污染羽识别的新方法, 能够以尽可能少的采样点达到污染羽准确识别的目的.

## 1 污染羽识别算法

在进行地下水污染治理和修复过程中, 通常要以尽可能少的水质采样点达到准确识别地下水污染羽的目的. 在场地信息不确定的前提下, 可以选择使污染浓度场的不确定度减少最大的点作为采样

点<sup>[11-12]</sup>.

### 1.1 基本假设

①污染源的强度已知;②地下水模型仅考虑渗透系数的不确定性.

### 1.2 算法流程

图 1 为污染羽识别算法的流程图. 具体计算步骤如下: ①确定污染源可能存在的位置, 并对每个可能位置赋以概率值作为初始权重. 权重值介于 0~1 之间, 0 表示该位置不存在污染源, 1 表示该位置存在污染源;②利用地质统计软件 SGeMS 生成场地的渗透系数随机场<sup>[13]</sup>;③构建地下水流模型与溶质运移模型, 通过蒙特卡罗 (Monte Carlo) 随机模拟得到污染浓度场库;④根据污染源权重随机选取污染浓度场, 提取滤波区的浓度场, 经统计分析得到初始的综合污染羽和浓度误差协方差矩阵;⑤结合已有采样点资料, 利用卡尔曼滤波技术对综合污染羽及误差协方差矩阵进行更新;⑥通过误差协方差矩阵寻找新采样点;⑦结合新采样点利用卡尔曼滤波技术更新综合污染羽;⑧利用模糊集合理论将综合污染羽与单个污染源所对应的污染羽进行对比, 得到污染羽的相似度;⑨根据相似度更新污染源权重. ⑩判断权重值和不确定度是否趋于稳定, 满足则停止

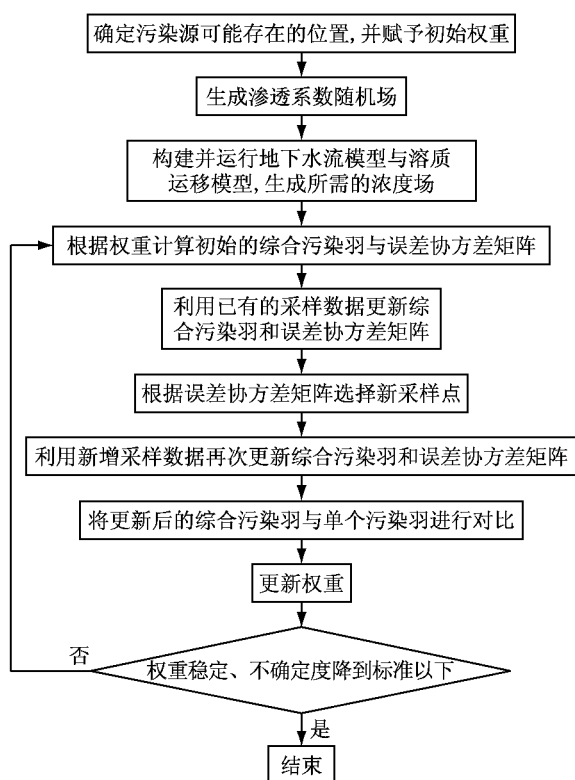


图 1 污染羽识别算法流程

Fig.1 Flow chart of the algorithm for identification of contaminant plume

计算, 否则回到步骤④.

### 1.3 综合污染羽生成与误差协方差估计

借鉴文献<sup>[14]</sup>的个体适应值库的概念, 在利用蒙特卡罗随机模拟方法生成污染浓度场时将模拟结果存入污染浓度场库. 如果污染源可能存在于  $m$  个位置, 且渗透系数随机场有  $q$  次实现, 则将可能的污染源与渗透系数场进行组合, 可生成  $m \times q$  个污染浓度场.

根据污染源权重大小, 从各个污染源所对应的  $q$  个浓度场 (对应于渗透系数场的  $q$  次实现) 分别随机选取浓度场. 在所有被选取出的污染浓度场中将相同渗透系数场所对应的浓度场叠加, 得到当前污染源权重下的浓度随机场的  $q$  次实现, 即统计意义上的  $q$  个样本.

对浓度随机场数据 ( $q$  次实现) 进行统计分析, 可以估计出综合污染羽的浓度值以及浓度误差协方差矩阵. 浓度随机场 ( $q$  次实现) 的均值即为综合污染羽的浓度, 节点  $j$  处的浓度值记为  $\bar{c}_j$ . 浓度误差协方差矩阵中的元素为  $R(c_i, c_j)$ , 其表达式为

$$R(c_i, c_j) = \frac{1}{q-1} \sum_{k=1}^q (c_{i,k} - \bar{c}_i)(c_{j,k} - \bar{c}_j) \quad (1)$$

式中:  $c_{i,k}$  与  $c_{j,k}$  分别表示在第  $k$  个污染浓度场中位置  $i$  和位置  $j$  处的浓度值;  $\bar{c}_i$  与  $\bar{c}_j$  分别表示在位置  $i$  和位置  $j$  处的平均浓度值 (即综合污染羽的浓度值). 需要说明的是, 上述对浓度随机场 ( $q$  次实现) 数据的统计分析 (综合污染羽生成和误差协方差矩阵估计) 并未在整个研究区 (含水层) 进行, 而只在滤波区 (即应用卡尔曼滤波技术的区域) 进行.

## 2 污染羽的更新和新采样点的选取

### 2.1 综合污染羽的更新

卡尔曼滤波是由卡尔曼在 1960 年推导出的一种线性、无偏且误差方差最小的优化估计方法, 它在已知系统模型和量测数学模型、量测噪声统计特征及系统状态的初始情况下利用量测数据获取系统状态变量的最优估计值<sup>[15]</sup>.

利用卡尔曼滤波技术通过连续作用采样点更新综合污染羽与误差协方差矩阵. 卡尔曼滤波的更新方程如下. 卡尔曼增益矩阵为

$$K = P_- \cdot H^T [H \cdot P_- \cdot H^T + r]^{-1} \quad (2)$$

浓度更新方程为

$$\hat{c}_+ = \hat{c}_- + K(z - H \cdot \hat{c}_-) \quad (3)$$

误差协方差更新方程为

$$\mathbf{P}_+ = [\mathbf{I} - \mathbf{K} \cdot \mathbf{H}] \cdot \mathbf{P}_- \quad (4)$$

式中:  $\mathbf{K}$  为卡尔曼增益矩阵;  $\mathbf{P}$  为  $n$  阶浓度误差协方差矩阵, 下标  $-$ ,  $+$  分别表示先验估计和后验估计, 对应利用测量数据更新之前和更新之后的状态;  $\mathbf{H}$  为量测矩阵, 通常为  $l \times n$  阶的矩阵,  $l$  为取样点的数目,  $n$  为用于计算的网格点个数(滤波区内);  $r$  为采样误差的方差;  $\hat{\mathbf{c}}$  是  $n$  维状态向量, 表示浓度估计值;  $z$  为采样值;  $\mathbf{I}$  为  $n$  阶单位矩阵. 通过连续添加单个采样点对综合污染羽及误差协方差矩阵进行更新, 此时  $\mathbf{H}$  为一个  $1 \times n$  阶的矩阵, 其形式如下:

$$\mathbf{H} = [0, 0, \dots, 0, 1, 0, \dots, 0] \quad (5)$$

其中, 向量元素仅在采样点处取 1, 其他位置取 0. 若已知采样点的位置和采样值, 可以利用式(2)~(5)更新综合污染羽. 需要说明的是, 初始的  $\hat{\mathbf{c}}_-$  和  $\mathbf{P}_-$  的计算方法见 1.3 节.

## 2.2 新采样点的选取

随着采样点的增加, 污染浓度场的不确定度(即误差协方差矩阵的对角线之和)会逐渐降低. 新采样点的选取方法如下: 首先将滤波区已有采样点之外的其他网格点视为备选采样点, 分别利用卡尔曼滤波的更新方程(式(4)和式(5))得到更新后的误差协方差矩阵, 然后计算各备选采样点对应的不确定度, 最后选择不确定度最小的点作为新采样点.

本文采用一种更为简便的方法, 就是寻找更新后的不确定度下降最大的点, 将其作为新采样点. 更新后的不确定度可以表示为更新前的不确定度减去新增采样点使不确定度下降的量<sup>[11]</sup>, 即

$$\sigma_{\text{T}}^2 = \sum_i P_{+,ii} = \sum_i P_{-,ii} - \frac{1}{P_{-,jj} + r_j} \sum_i P_{-,ij}^2 \quad (6)$$

式中:  $\sigma_{\text{T}}^2$  为总方差, 即协方差矩阵的对角线元素之和;  $P_{+,ii}$  为矩阵  $\mathbf{P}_+$  的元素  $(i, i)$ ;  $\sum_i P_{-,ii}$  为更新前的不确定度;  $\sum_i P_{+,ii}$  为更新后的不确定度;  $r_j$  为位置  $j$  处采样误差的方差. 为了使不确定度达到最小, 采样点应选在使  $\frac{1}{P_{-,jj} + r_j} \sum_i P_{-,ij}^2$  最大的位置.

## 3 污染羽的模糊对比

单个污染源的污染羽定义为该污染源在渗透系数随机场的所有实现下的浓度均值. 如果有 5 个污染源, 则可以生成 5 个单个污染源的污染羽. 通过将单个污染源的污染羽与利用卡尔曼滤波更新后的综合污染羽进行对比, 对污染源可能所在位置的权重

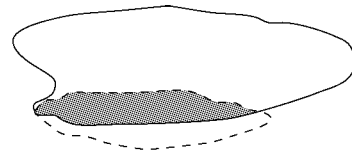
进行更新, 从而确定污染源位置.

首先将污染羽用模糊集来表示, 隶属函数为标准化的浓度值  $c'$  (定义为浓度值除以最大浓度值). 在考虑数个模糊集时, 每个模糊集中的值均小于一个给定的标准  $\alpha_s$ . 若  $\alpha_s$  分别取值为 0.2, 0.4, 0.6, 0.8, 则共可得到 4 个模糊集.

在某一  $\alpha_s$  下, 将单一污染源的污染羽的模糊集分别与综合污染羽的模糊集进行对比, 计算交叉面积为  $S_s$ , 如图 2 所示的交叉面积表征单一污染源的污染羽与综合污染羽的相似度. 全局相似度  $g$  为各个标准  $\alpha_s$  下的相似度的加权和<sup>[12]</sup>, 即

$$g = \sum_s \alpha_s S_s \quad s = 1, 2, 3, \dots \quad (7)$$

图 2 为 2 个模糊集的比较, 它一方面考虑了污染羽截面的形状, 另一方面也考虑对污染羽中浓度较高的区域赋以较大的权重. 污染源新的权重通过污染源全局相似度进行标准化后得到, 具体就是将各相似度除以它们中的最大值, 得到的就是标准化的相似度.



— 综合污染羽 - - - 单个污染源的污染羽

图 2 综合污染羽与单个污染源污染的模糊比较

Fig.2 Comparison of composite plume and individual source plume

## 4 算例

### 4.1 问题概述

如图 3 所示, 研究区存在二维非均质各向同性

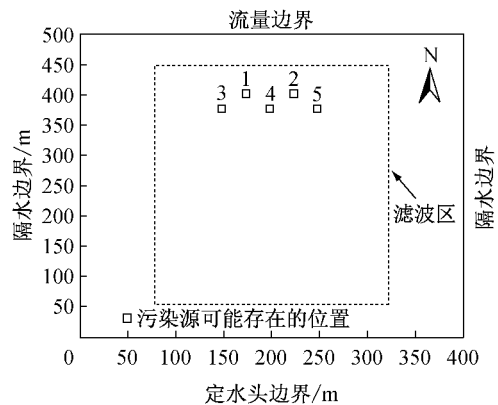


图 3 含水层结构平面

Fig.3 Plan view of confined aquifer

的承压含水层,尺寸为  $400\text{ m} \times 500\text{ m}$ ,用边长为  $5\text{ m}$  的正方形网格将研究区域剖分为 100 行、80 列的有限差分网格.假设含水层水流运动为稳定流,含水层厚度为  $10\text{ m}$ .上边界为二类边界,单宽流量为  $1.0\text{ m}^3 \cdot \text{d}^{-1}$ ,下边界为定水头边界,水头为  $0\text{ m}$ ,左、右边界为隔水边界.含水介质孔隙度为  $0.3$ ,纵向弥散度为  $40.0\text{ m}$ ,水平横向弥散度为  $1.2\text{ m}$ .已知污染源体积注入率和污染物浓度,分别为  $1\text{ m}^3 \cdot \text{d}^{-1}$  和  $500\text{ mg} \cdot \text{L}^{-1}$ ,但污染源位置未知.

含水层的渗透系数为一随机变量,且服从对数正态分布,其均值  $\ln K$  和标准差  $\sigma_{\ln K}$  分别为  $2.62$  和  $0.53$ ,在  $x$  方向和  $y$  方向的相关长度均为  $140\text{ m}$ ,变差函数为指数型.图 4 为渗透系数  $\ln K$  场的一次实现.根据已有的场地信息,污染源可能存在的位置如图 5 所示,其中位置 2 为污染源的真正位置(事先未知).本算例是在含水层参数不确定、污染源位置未知条件下通过择优选择采样点识别地下水污染羽.

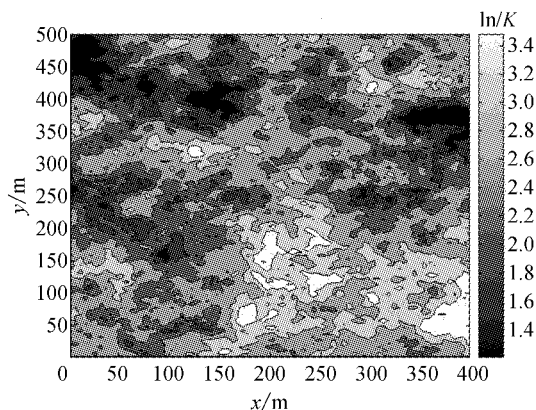


图 4 渗透系数随机场的任意一次实现  
Fig.4 One randomly selected stochastic hydraulic conductivity field

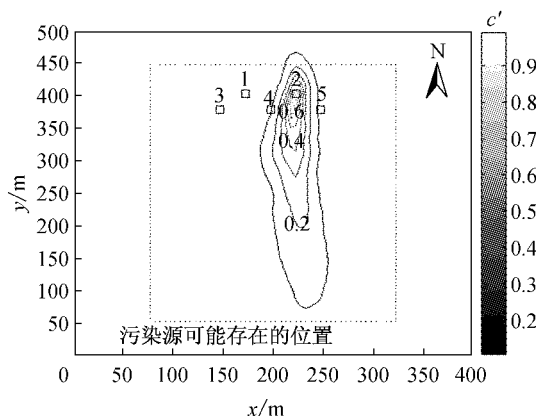


图 5 真实的污染源位置和对应的污染羽分布  
Fig.5 True source location and the corresponding plume distribution

本算例中,渗透系数随机场共有 100 次实现,污染源可能存在的位置为 5 处,将渗透系数随机场的每次实现分别与各个可能的污染源结合,进行蒙特卡罗随机模拟共得到 500 个污染浓度场.为了减少调用模拟计算的次数,将这 500 个污染浓度场存入污染浓度场库.之后的计算中,只需要根据每次更新后的污染源权重随机选取浓度场即可.卡尔曼滤波区为图 3 虚线包围的网格.

污染羽识别算法的主要参数设置如下:污染源初始权重分别设为  $0.80, 0.85, 0.90, 0.95, 1.00$ .污染羽的模糊对比采用 3 个模糊集,对应的标准为  $0.2, 0.4$  和  $0.6$ .

## 4.2 算例分析

图 6a 为根据污染源初始权重计算得到的综合污染羽,图 6c—6h 为利用 1 个、6 个和 7 个采样点更新后的污染羽及其对应的污染源权重.

由图 6c—6h 可知,经过 7 次采样后,位置 2 权重稳定于 1,其他位置权重接近于零,同时将图 6g 与真实的污染羽形态进行比较,两者形态十分接近,表明只需要 7 个采样数据就可以确定污染源的位置,并能够达到较为准确地识别污染羽的目的.此外,由图 6g 可以看出,4 个采样点(点 2, 4, 5, 6)分布在临近污染源的下流方向,其余 3 个采样点(点 1, 3, 6)起到控制污染羽形态的作用.上述结果表明,采样点的分布主要考虑 2 个因素,一是距离污染源的位置,二是污染浓度场的不确定度.

图 7 为浓度场的不确定度随采样点数目的变化情况.从图中可以看出,随着采样点数目的增加,浓度场的不确定度随之降低,并且当采样点的数目增加至 7 时,浓度场的不确定度降到了较低的水平,为  $237.45\text{ (mg} \cdot \text{L}^{-1})^2$ ,即平均标准差为  $0.24\text{ mg} \cdot \text{L}^{-1}$ ,已达到污染羽识别的精度要求,结束计算.

## 5 结论

(1)基于卡尔曼滤波技术和模糊集合理论提出一类地下水污染羽识别的方法,算例研究表明该方法是一种有效的污染羽识别算法,能够以尽可能少的采样点确定污染源的位置,并较为准确地达到识别污染羽的目的.

(2)该方法应用了污染浓度场库的概念,只需要在计算开始时利用数值模型生成多个浓度场,此后的计算中不需要重新计算浓度场,节省了计算成本.

(3)所建立的污染羽识别方法目前只是考虑了

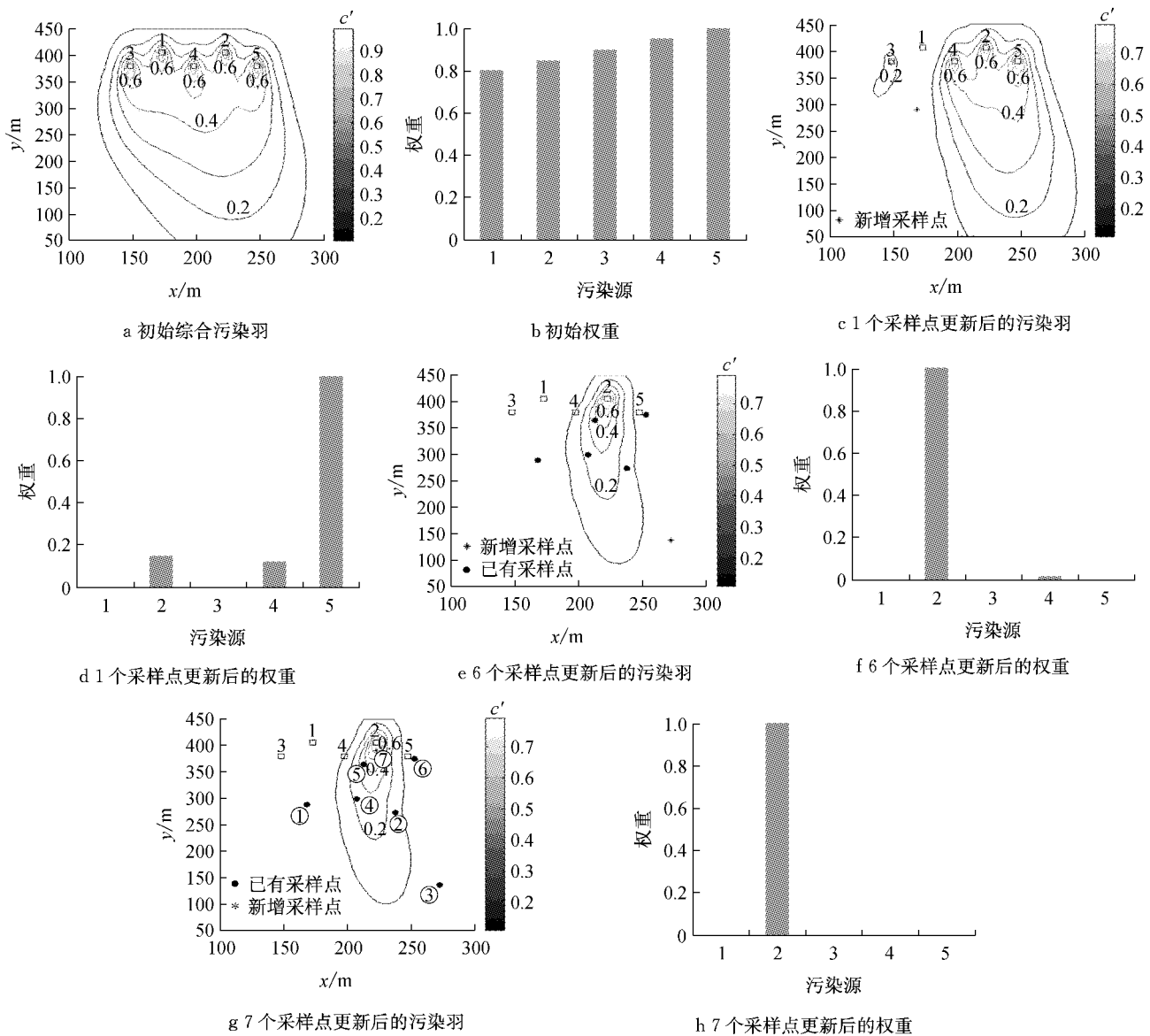


图6 综合污染羽与污染源权重随采样的变化

Fig.6 The change of composite plume and its source location weights with sampling

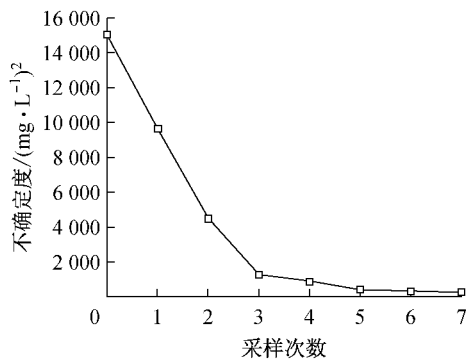


Fig.7 Convergence of uncertainty with the procedure of sampling

渗透系数的不确定性,并且污染源的强度假设是已知的.在下一步的研究中要考虑其他水文地质参数

的不确定性影响下污染源的位置和强度的同步识别.

#### 参考文献:

- [1] 吴剑锋,郑春苗. 地下水污染监测网的设计研究进展[J]. 地球科学进展, 2004, 19(3): 429.  
WU Jianfeng, ZHENG Chunmiao. Contaminant monitoring network design: recent advances and future directions [J]. Advance in Earth Sciences, 2004, 19(3):429.
- [2] Chien C C, M A Medina, G F Pinder, et al. Environmental modeling and management: theory, practice and future directions [M]. [S.l.]: Today Media Inc, 2002.
- [3] Wu J, Zheng C, Chen C C. Cost-effective sampling network design for contaminant plume monitoring under general hydrogeological conditions [J]. Journal of Contaminant Hydrology, 2005, 77(1/2):41.

- [4] Dhar A, Datta B. Global optimal design of ground water monitoring network using embedded Kriging [J]. *Ground Water*, 2009, 47(6): 806.
- [5] Owlia R R, Abrishamchi A, M Tajrishy. Spatial-temporal assessment and redesign of groundwater quality monitoring network: a case study [J]. *Environmental Monitoring and Assessment*, 2011, 172(1/4): 263.
- [6] Datta B, Chakrabarty D, Dhar A. Optimal dynamic monitoring network design and identification of unknown groundwater pollution sources [J]. *Water Resources Management*, 2009, 23(10): 2031.
- [7] Kim K, Lee K. Optimization of groundwater-monitoring networks for identification of the distribution of a contaminant plume [J]. *Stochastic Environmental Research and Risk Assessment*, 2007, 21(6): 785.
- [8] 王开丽, 黄冠华. 二维强非均质含水层中渗透系数空间变异对污染物迁移的影响[J]. *水利学报*, 2010, 41(4): 437.  
WANG Kaili, HUANG Guanhua. Effect of hydraulic conductivity on contaminant transport in highly heterogeneous two-dimensional aquifers [J]. *Journal of Hydraulic Engineering*, 2010, 41(4): 437.
- [9] 阎婷婷, 吴剑锋. 渗透系数的空间变异性对污染物迁移的影响研究[J]. *水科学进展*, 2006, 17(1): 29.  
YAN Tingting, WU Jianfeng. Impacts of the spatial variation of hydraulic conductivity on the transport fate of contaminant plume [J]. *Advances in Water Science*, 2006, 17(1): 29.
- [10] 施小清, 吴吉春, 吴剑锋, 等. 多个相关随机参数的空间变异性对溶质运移的影响[J]. *水科学进展*, 2012, 23(4): 509.  
SHI Xiaoqing, WU Jichun, WU Jianfeng, et al. Effects of the heterogeneity of multiple correlated random parameters on solute transport [J]. *Advances in Water Science*, 2012, 23(4): 509.
- [11] Herrera G S, Pinder F G. Space-time optimization of groundwater quality sampling networks [J]. *Water Resources Research*, 2005, 41(12): 407.
- [12] Dokou Z, Pinder F G. Optimal search strategy for the definition of a DNAPL source[J]. *Journal of Hydrology*, 2009, 376: 542.
- [13] Remy N, Boucher A, Wu B J. Applied geostatistics with SGeMS: a user's guide[M]. New York: Cambridge University Press, 2009.
- [14] 吴剑锋, 彭伟, 钱家忠, 等. 基于 INPGA 的地下水污染治理多目标优化管理模型: I——理论方法与算例验证[J]. *地质论评*, 2011, 57(2): 277.  
WU Jianfeng, PENG Wei, QIAN Jiazhong, et al. INPGA-based multi-objective management model for optimal design of groundwater remediation system: I—methodology and its experimental validation [J]. *Geological Review*, 2011, 2: 277.
- [15] 刘胜, 张红梅. 最优估计理论[M]. 北京: 科学出版社, 2011.  
LIU Sheng, ZHANG Hongmei. Theory of optimal estimation [M]. Beijing: Science Press, 2011.

### (上接第 434 页)

- ZHANG Shuai. The optimal design study of urban rail transit hub internal space traffic streamline [D]. Beijing: Beijing Jiaotong University, 2011.
- [11] 马岳. 轨道交通枢纽个体乘客流线组织研究[D]. 西安: 长安大学, 2011.  
MA Yue. Passenger rail transportation hub streamline organization of the individual [D]. Xi'an: Chang'an University, 2011.
- [12] 刘莹. 基于乘客行为特征的轨道站客流路径选择研究[D]. 上海: 同济大学, 2009.  
LIU Ying. Research on route choice in rail transit station based on pedestrian travel behavior [D]. Shanghai: Tongji University, 2009.
- [13] 胡春平, 毛保华, 朱宇婷. 综合客运枢纽旅客全过程流线优化模型[J]. *交通运输工程系统与信息*, 2012, 12(3): 159.  
HU Chunping, MAO Baohua, ZHU Yuting. Streamline design and optimization for whole process of passenger transfer in integrated transportation hubs [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2012, 12(3): 159.
- [14] 赵莉. 城市轨道交通枢纽交通设计理论与方法研究[D]. 北京: 北京交通大学, 2011.  
ZHAO Li. Research on traffic design theory and methods of urban rail transit hubs [D]. Beijing: Beijing Jiaotong University, 2011.
- [15] 唐明. 客运枢纽行人交通行为模型与仿真算法研究[D]. 吉林: 吉林大学, 2010.  
TANG Ming. Research on pedestrian traffic behavior model and simulation algorithm in passenger transfer hub [D]. Changchun: Jinlin University, 2010.
- [16] 李岩. 基于路径分配的城市轨道交通枢纽疏散管理研究[D]. 北京: 北京交通大学, 2010.  
LI Yan. Research on evacuation management of urban rail transportation center based on path assignment [D]. Beijing: Beijing Jiaotong University, 2010.
- [17] 张晓军, 张宁, 陈晖. 城市轨道交通行人通道阻抗函数拟合[J]. *城市轨道交通研究*, 2011, 14(1): 33.  
ZHANG Xiaojun, ZHANG Ning, CHEN Hui. Impedance function's fitting of passenger corridor in urban rail transit [J]. *Urban Mass Transit*, 2011, 14(1): 33.
- [18] 漆凯. 城市客运枢纽站旅客流线优化研究[D]. 北京: 北京交通大学, 2012.  
QI Kai. Research on optimization for flow routing of urban passenger hubs [D]. Beijing: Beijing Jiaotong University, 2012.
- [19] 于学军. 城市轨道交通枢纽流线费用模型研究[D]. 北京: 北京交通大学, 2012.  
YU Xuejun. Research on flow routing cost model of urban rail transit hub [D]. Beijing: Beijing Jiaotong University, 2012.
- [20] 李三兵. 城市轨道交通车站客流特征与服务设施的关系研究[D]. 北京: 北京交通大学, 2009.  
LI Sanbing. Study on the relationship of passenger flow and service facilities in urban rail transit station [D]. Beijing: Beijing Jiaotong University, 2009.
- [21] National Research Council. Highway capacity manual 2000[R]. Washington D C: TRB, 2000.