

DOI:10.3876/j.issn.1000-1980.2018.05.005

基于改进型遗传算法的水质溯源问题

李敏慧¹,荆立¹,朱彦君²,郢新宇³,王青¹,孔俊¹

(1. 河海大学江苏省海岸海洋资源开发与环境安全重点实验室,江苏南京 210098;
2. 中设设计集团股份有限公司,江苏南京 210014; 3. 南京信息工程大学数学与统计学院,江苏南京 210044)

摘要: 针对污染源排放位置、时间、排放总量的问题,以一维物质对流扩散方程和遗传算法理论为基础,构建污染物位置和排放时间的溯源模型。在传统遗传算法的基础上,按照适应度函数值的大小进行排序,引入“权重系数法”的概念,生成产生子代的新规则,优化种群质量,以控制种群进化方向。进一步基于解耦技术,建立了污染源排放总量的优化模型,推求污染源释放总量。试验结果表明,改进后的遗传算法能准确地追踪到污染源的排放位置和排放时间,与传统方法相比,其搜索效率更高,计算更加稳定。

关键词: 污染源排放;水质溯源;遗传算法;优化模型;参数解耦;权重系数法

中图分类号: X522 **文献标志码:** A **文章编号:** 1000-1980(2018)05-0402-06

Research on the water source identification based on an improved genetic algorithm

LI Minhui¹, JING Li¹, ZHU Yanjun², GAO Xinyu³, WANG Qing¹, KONG Jun¹

(1. Jiangsu Key Laboratory of Coast Ocean Resources Development and Environment Security, Hohai University, Nanjing 210098, China;

2. China Design Group Co., Ltd., Nanjing 210014, China;

3. School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: In order to solve the tracing problem of pollutant source in the subject of water environment, a backtracking model was built in this study to discern the source location and the emission time, based on the convection-diffusion equation and the genetic algorithm. By further introducing the decoupling method, the mass quality was also obtained. A concept of ‘weigh coefficient’ was introduced to improve the traditional genetic algorithm for high computational efficiency by controlling the evolutionary trend of population and optimizing the quality of population. Compared with the traditional genetic algorithm, the improved one makes the backtracking model more efficient and stable.

Key words: pollution source emission; water source identification; genetic algorithm; optimized model; parameter decoupling; weigh coefficient method

水污染溯源问题是分析污染源位置、排放时刻和浓度的一类水力学反问题^[1-2],具有非线性和不稳定性^[3-5]。目前,国内外水污染溯源研究仍处于探索阶段。辛小康等^[6]将遗传算法与数学分析法结合,建立的污染源模型能识别单点源和多点源水质污染源;刘晓东等^[7]提出了以有限差分法为正问题求解器研究一维污染源识别问题的方法;牟行洋等^[8]采用微分进化算法进行二维环境水力学模型的参数反演;Jha等^[9]采用

基金项目: 水利部公益性行业科研专项(201501010)

作者简介: 李敏慧(1996—),女,硕士研究生,主要从事水质溯源研究。E-mail:513914101@qq.com

通信作者: 孔俊,教授。E-mail:kongjun999@126.com

引用本文: 李敏慧,荆立,朱彦君,等. 基于改进型遗传算法的水质溯源问题[J]. 河海大学学报(自然科学版),2018,46(5):402-407.

LI Minhui, JING Li, ZHU Yanjun, et al. Research on the water source identification based on an improved genetic algorithm[J]. Journal of Hohai University(Natural Sciences), 2018, 46(5): 402-407.

自适应模拟退火算法对地下水污染进行溯源研究;陈媛华等^[10]建立了线性回归模型,由回归分析求解排放源的排放位置、时间、强度;Wei 等^[11]设计了一种基于最佳摄动量正则化的耦合方法,对多点源分数阶扩散方程进行溯源分析;王家彪等^[12]构建了基于耦合概率密度方法的溯源模型,利用微分进化算法实现模型求解;杨海东等^[13]提出微分进化蒙特卡罗思想,在贝叶斯-蒙特卡罗方法的基础上,结合微分进化的思想,并将之运用到二维河道瞬时突发水污染溯源研究中。

鉴于遗传算法寻优能力和适应性强的优势,在此基础上采用正向模拟以逼近实测浓度分布的思路被广泛采纳^[14]。但传统遗传算法的收敛精度和种群质量不高常导致溯源低效,为此,本文侧重对遗传算法进行改进。

1 污染溯源问题描述

河道中可溶性污染物的输运过程,可采用一维对流扩散方程描述^[15-16]:

$$\frac{\partial C(x,t)}{\partial t} + \frac{\partial (uC(x,t))}{\partial x} = E \frac{\partial^2 C(x,t)}{\partial^2 x} - kC(x,t) \tag{1}$$

式中: $C(x,t)$ ——河渠 x 断面 t 时刻污染物的平均质量浓度,mg/L; u ——河渠断面平均流速,m/s; E ——河渠纵向离散系数,m²/s; k ——污染物降解系数,s⁻¹,质量守恒情况下,不考虑污染物的降解, k 取 0。

对应恒定流情况,方程(1)的解析解为^[17]

$$C(x,t) = \frac{m_0}{\sqrt{4\pi E(t-t_0)}} \exp\left\{-\frac{[x-x_0-u(t-t_0)]^2}{4E(t-t_0)} - k(t-t_0)\right\} \tag{2}$$

式中: x_0 ——污染源位置,m; t_0 ——污染发生时间,s; m_0 ——污染物初始面源强度,g/m²。

针对溯源问题,需结合污染物浓度的时空分布推求式(2)中的 x_0 、 t_0 、 m_0 。目前的求解方法主要基于正向浓度场预测浓度最优化的思路,结合遗传算法不断进行试算,使计算浓度不断逼近观测浓度^[12]。

2 遗传算法的优化

先假定 x_0 、 t_0 、 m_0 的范围,利用遗传算法产生 x_0 、 t_0 的初始代及子代,当子代最接近污染源参数真实值时,计算浓度与观测浓度的误差最小。传统遗传算法随机性强,无法控制子代总是向优良方向进化。为提高算法效率,笔者提出“权重系数法”来改进遗传算法。权重系数法,即对传统遗传算法产生的中间个体按适应度函数值升序排序,即按质量从好到坏排列。选取个体时,前 $N/2$ 个个体不变,后 $N/2$ 个个体与前 $N/2$ 个个体按照一定权重组合,生成新个体。

a. 种群初始化。给定种群规模 N 和边界约束条件,由式(3)生成第一代个体 $X_{i,\text{Gen}}(x_0, t_0)$ ($i=1, \dots, N$):

$$X_{i,\text{Gen}} = X_{\min} + A(X_{\max} - X_{\min}) \tag{3}$$

式中: A —— $[0,1]$ 的随机数。

引入适应度函数 F 以最小化为最优:

$$F(X_i) = |1 - r(X_i)| \tag{4}$$

式中: r ——相关系数,其表达式见文献[12]。

目标函数为

$$\min F(X_i) \tag{5}$$

优化模型(4)(5)可实现污染源强度、位置和时间参数的解耦^[12],通过先确定位置和时间参数后再确定污染物面源强度 m_0 ,文献[18]中给出了 m_0 的求解公式。

b. 变异。由传统遗传算法产生中间变异个体 $V_{zi,\text{Gen}}$,产生方式如式(6),若 $V_{zi,\text{Gen}} \notin [X_{\min}, X_{\max}]$,按式(6)重新生成变异个体 $V_{zi,\text{Gen}}$,将其按 $F(V_{zi,\text{Gen}})$ 的值升序排序。按式(7)产生改进后的个体 $V_{i,\text{Gen}}$

$$V_{zi,\text{Gen}} = X_{r_1,\text{Gen}} + P_m(X_{r_3,\text{Gen}} - X_{r_2,\text{Gen}}) \tag{6}$$

$$V_{i,\text{Gen}} = \begin{cases} V_{zi,\text{Gen}} & i = 1, N/2 \\ \alpha V_{zi,\text{Gen}} + (1 - \alpha) V_{z(N-i+1),\text{Gen}} & i = N/2 + 1, N \end{cases} \tag{7}$$

式中: P_m ——变异概率,通常取 $[0.5,1]^{[19]}$; α ——权重系数。 $r_1, r_2, r_3=1, \cdots, N$ 且 r_1, r_2, r_3 均不为 i 。

c. 交叉。由传统遗传算法按式(8)产生中间交叉个体 $U_{zi, Gen}$,将其按 $F(U_{zi, Gen})$ 值升序排序,按式(9)产生交叉个体 $U_{i, Gen}$ 。

$$U_{zi, Gen}(x_j)=\begin{cases} V_{i, Gen}(x_j) & (0,1) < P_c \text{ 或 } j=j_r \\ X_{i, Gen}(x_j) & \text{其他} \end{cases} \tag{8}$$

$$U_{i, Gen}=\begin{cases} U_{zi, Gen} & i=1, N/2 \\ \alpha U_{zi, Gen}+(1-\alpha)U_{z(N-i+1), Gen} & i=N/2+1, N \end{cases} \tag{9}$$

式中: P_c ——交叉概率,常取 $[0.8,1]^{[19]}$; j_r ——1,2 中随机取值; x_j —— X_i 的第 j 个分量。

d. 选择。适应度函数值越小,表示该个体适应度越好,由传统遗传算法按式(10)产生中间选择个体 $X_{zi, Gen+1}$,将其按 $F(X_{zi, Gen+1})$ 的值升序排序。改进后的遗传算法中,选择个体 $X_{i, Gen+1}$ 产生方式如式(11)。

$$X_{i, Gen+1}=\begin{cases} U_{i, Gen} & F(U_{i, Gen}) < F(X_{i, Gen}) \\ X_{i, Gen} & F(X_{i, Gen}) < F(U_{i, Gen}) \end{cases} \tag{10}$$

$$X_{i, Gen+1}=\begin{cases} X_{zi, Gen+1} & i=1, N/2 \\ \alpha X_{zi, Gen+1}+(1-\alpha)X_{z(N-i+1), Gen+1} & i=N/2+1, N \end{cases} \tag{11}$$

由式(11)产生最后一代新个体中适应度最好的个体,其代表的参数即为所求 x_0, t_0 ,再根据文献[18]给的公式求出 m_0 ,代入式(2)的溯源模型中,探求污染过程。

3 算例分析

3.1 算例1

设有一河道,底宽10.0m,边坡2.0,水深3.5m,底坡1/10000,曼宁糙率0.017,离散系数为2.0m²/s,流量为15.0m³/s。假定某日上午9:00有800.0kg 污染物排入0m处,在1000m和2000m处设监测点,9:30开始监测,数据见表1。参数范围为 $[(-500\text{ m}, 0\text{ s}), (1000\text{ m}, 7200\text{ s})]$,种群数 N 为50。

3.1.1 权重系数α的确定

改进算法中引入的权重系数直接影响改进后遗传算法的收敛效率。理论上,α 越小,收敛速度越快。从表2可以看出,α 在 $[1/10, 1/3]$ 之间取值时,收敛效率相差不大,且适应度函数值变化不大。为了保证种群的多样性,α 建议取 $[1/10, 1/3]$,本文选取 $\alpha=1/10$ 。

3.1.2 结果分析及比较

3.1.2.1 稳定性分析

如图1所示,采用均方差 $\sigma^2(F) = \left[\sum (F(X_{i, Gen}) - F_{\min})^2 \right] / N$ 评价收敛速度快慢,其值越大则种群散度较大(其中, $F_{\min}$ 为 x_0, t_0, m_0 取真实值时的适应度函数, $F_{\min}=0$)。从模拟结果可知,迭代次数小于10时,改进遗传算法中,适应度函数方差总是小于传统遗传算法适应度函数的方差。这表明与传统遗传算法相比,改进遗传算法能更稳定地收敛于真实值附近,有助于快速确定污染源的相关参数范围。

3.1.2.2 收敛速度比较

假设7:30时 $t=0$,那么9:00排放污染物, t_0 的真实值为5400s。由图2可知,传统遗传算法在本例中迭代20次时,

表1 实验观测断面污染物质量浓度过程
Table 1 Observed concentration data of the test section

时刻	$\rho(C_1)/(\text{mg} \cdot \text{L}^{-1})$	时刻	$\rho(C_2)/(\text{mg} \cdot \text{L}^{-1})$
9:30	0	10:30	0
9:50	4.125	11:00	17.478
10:00	33.222	11:10	29.834
10:10	37.321	11:30	10.359
10:30	1.771	12:00	0.060
11:00	0	12:30	0

注: $\rho(C_1), \rho(C_2)$ 分别为1000m,2000m处污染物质量浓度。

表2 α 与 n、F(X) 的关系
Table 2 Relationship between α, n and F(X)

α	n	F(X)/10 ⁻⁴	α	n	F(X)/10 ⁻⁴
1/2	20	2.51	1/7	12	2.51
1/3	14	2.52	1/8	14	2.51
1/4	11	2.52	1/9	14	2.51
1/5	12	2.52	1/10	10	2.51
1/6	10	2.51			

注:n 为迭代次数。

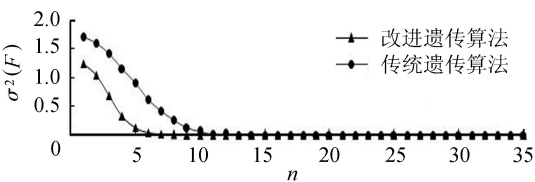


图1 适应度函数的方差随迭代次数变化情况
Fig.1 Relationship between $\sigma^2(F)$ and iteration

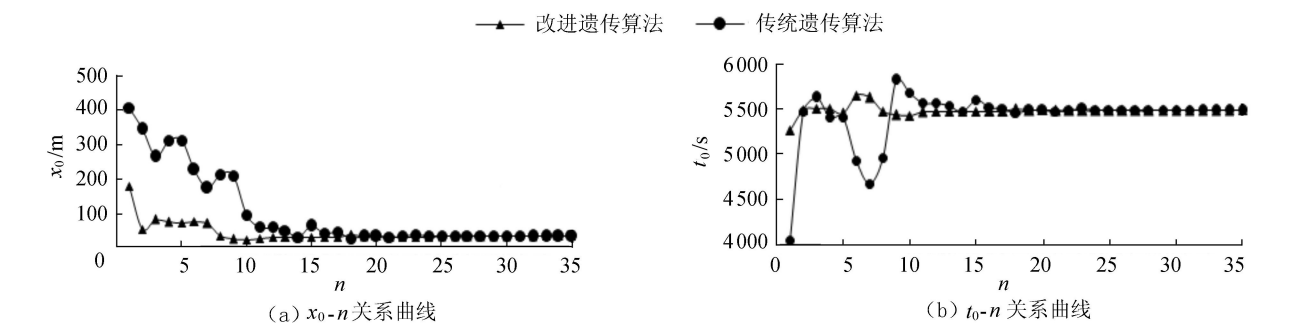


图 2 迭代过程曲线

Fig. 2 Comparison between the iterative curves of model parameters

x_0 、 t_0 趋于稳定,而改进遗传算法只需迭代 10 次, x_0 、 t_0 便趋于稳定。可见传统遗传算法迭代十几次才能确定的范围,改进后的遗传算法只需迭代两三次便能做到。当给定的边界范围增大时,改进后的遗传算法能大幅度减小收敛时的迭代次数,提高搜索效率。

3. 1. 2. 3 计算精确分析

基于传统遗传算法和改进后的遗传算法,分别迭代 50 次,其溯源结果见表 3。对比文献[12]基于耦合概率密度给出本问题的溯源结果,由表 3 可以看出,改进后的遗传算法能准确地搜索到污染源的位置、时间和质量。

3. 2 算例 2

以长江中下游污染物溯源为例。图 3 为长江(南京段)地形及模拟投放点、测点分布图。在模拟投放点倾倒 36000 kg 污染物,从观测点监测到污染物开始,每 5 min 观测 1 次,连续观测 12 h。污染源位置在图 4 中以红色五角星标出,圆圈和十字分别表示改进前后生成的个体。图 4(a)(b)(c)分别为初始种群个体分布、迭代 5 次和迭代 10 次后种群个体分布。

表 3 溯源结果与真值比较

Table 3 Comparison between identified results and true values

统计类型	$m_0/$ ($\text{kg} \cdot \text{m}^{-2}$)	x_0/m	t_0/s
实测值	800	0	32400
传统遗传算法推算值	798.94	32.27	32488
耦合概率密度方法推算值	828.68	-70.77	32160
改进后的遗传算法推算值	798.96	31.91	32486

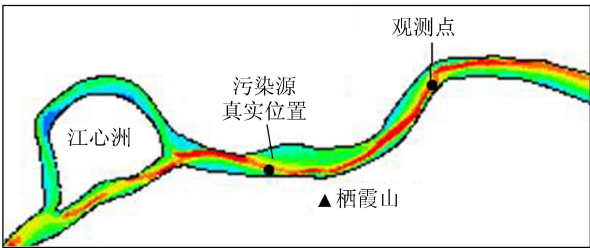


图 3 长江(南京段)测点分布

Fig. 3 Sketch of observing location in the Yangtze River

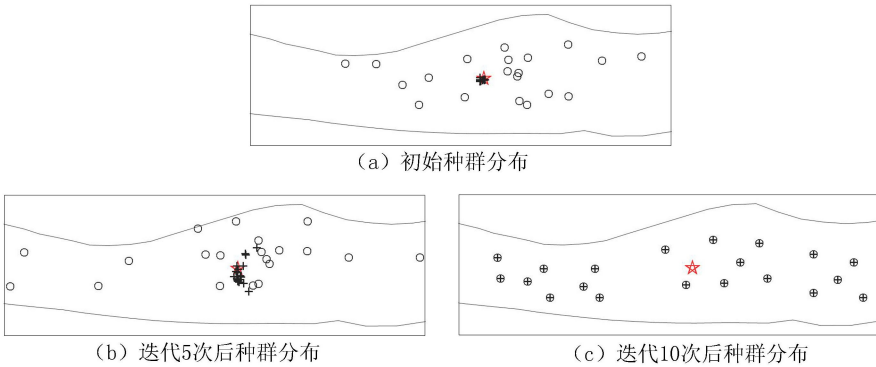


图 4 种群分布随迭代次数的变化

Fig. 4 Population distribution with iterations

由图4(b)(c)可知:改进前产生的个体分布散乱;改进后产生的个体均比改进前产生的个体距污染源真实值距离近,且迭代10次后,算法已经收敛,程序计算点逐渐集中到污染源的真正位置。对比基于传统遗传算法和改进遗传算法的结果,可见改进后的遗传算法收敛速度快,且适用于污染源溯源实例,详见表4。

表4 溯源结果与真值比较
Table 4 Comparison between identified results and true values

统计类型	x_0/m	y_0/m	统计类型	x_0/m	y_0/m
实测值	116 967. 0	564 115. 0	传统算法迭代10次推算值	120 389. 4	564 590. 2
传统算法迭代5次推算值	112 022. 9	563 732. 0	改进算法迭代10次推算值	116 884. 4	564 055. 5
改进算法迭代5次推算值	117 384. 4	564 567. 8	传统算法绝对误差	3 422. 4	475. 2
传统算法绝对误差	4 944. 1	383. 0	改进后算法绝对误差	82. 6	59. 5
改进后算法绝对误差	417. 4	452. 8			

注:(x_0,y_0)为污染源坐标。

4 结 论

针对溯源过程中采用的遗传算法存在的低效率问题,引入“权重系数法”,实现新生种群向最佳逼近方向的演化,进一步优化了次代种群的质量,提高了溯源效率。实例应用结果表明,改进后的遗传算法能更快地确定污染物发生的时间、位置及排放总量,搜索效率得到明显提高,预测结果精度能够得到保证。本文方法可为今后河口及海岸带的水质溯源问题提供一定的参考。

参考文献:

[1] 刘晓东,姚琪,薛红琴,等. 环境水力学反问题研究进展[J]. 水科学进展,2009,20(6):885-893. (LIU Xiaodong,YAO Qi,XUE Hongqin,et al. Advance in inverse problems of environmental hydraulics[J]. Advances in Water Science,2009,20(6):885-893. (in Chinese))

[2] 于英鹏,刘敏. 太湖流域水源地多环芳烃分布、溯源与生态风险评估[J]. 水资源保护,2017,33(3):82-89. (YU Yingpeng,LIU Min. Distribution,pollution sources,and ecological risk assessment of PAHs in water source area of Taihu Basin[J]. Water Resources Protection,2017,33(3):82-89. (in Chinese))

[3] HADAMARD J. Lectures on cauchy's problem in linear partial differential equations[M]. Chicago:Courier Corporation,2014.

[4] 金文龙,彭辉,高俊杰,等. 中小河道突发事故污染源强非稳态过程反问题[J]. 水资源保护,2013,29(2):18-21. (JIN Wenlong,PENG Hui,GAO Junjie,et al. Inverse problem of unsteady source intensity change process in sudden pollution accidents in medium-and small-sized rivers[J]. Water Resources Protection,2013,29(2):18-21. (in Chinese))

[5] 路川藤,黄华聪,钱明霞. 长江口北槽丁坝田区潮流及污染物迁移扩散特征[J]. 河海大学学报(自然科学版),2016,44(3):265-271. (LU Chuanteng,HUANG Huacong,QIAN Mingxia. Tidal current and pollutant diffusion characteristics in spur dike field of North Passage in Yangtze Estuary[J]. Journal of Hohai University (Natural Sciences),2016,44(3):265-271. (in Chinese))

[6] 辛小康,韩小波,李建,等. 基于遗传算法的水污染事故污染源识别模型[J]. 水电能源科学,2014,32(6):53-55. (XIN Xiaokang,HAN Xiaobo,LI Jian,et al. Pollutant source identification model of water pollution incident based on Genetic algorithm[J]. Water Resources and Power,2014,32(6):53-55. (in Chinese))

[7] 刘晓东,陈立强,华祖林,等. 河道一维污染源识别反问题[J]. 水力发电学报,2014,33(6):133-135. (LIU Xiaodong,CHEN Liqiang,HUA Zulin,et al. Inverse problem on pollutant source identification in 1D channel[J]. Journal of Hydroelectric Engineering,2014,33(6):133-135. (in Chinese))

[8] 牟行洋. 基于微分进化算法的污染源项识别反问题研究[J]. 水动力学研究与进展(A辑),2011,26(1):24-30. (MOU Xingyang. Research on inverse problem of pollution source term identification based on differential evolution algorithm[J]. Journal of Hydrodynamics(Ser A),2011,26(1):24-30. (in Chinese))

[9] JHA M,DATTA B. Three-dimensional groundwater contamination source identification using adaptive simulated annealing[J]. Journal of Hydrologic Engineering,2012,18(3):307-317.

[10] 陈媛华,王鹏,姜继平,等. 基于相关系数优化法的河流突发污染源项识别[J]. 中国环境科学,2011,31(11):1802-1807. (CHEN Yuanhua,WANG Peng,JIANG Jiping,et al. Contaminant point source identification of rivers chemical spills based on correlation coefficients optimization method[J]. China Environmental Science,2011,31(11):1802-1807. (in Chinese))

[11] WEI Hui,CHEN Wen,SUN Hongguang,et al. A coupled method for inverse source problem of spatial fractional anomalous

diffusion equations[J]. Inverse Problems in Science and Engineering,2010,18(7): 945-956.

[12] 王家彪,雷晓辉,廖卫红,等. 基于耦合概率密度方法的河渠突发水污染溯源问题[J]. 水利学报,2015,46(11):1280-1289. (WANG Jiabiao, LEI Xiaohui, LIAO Weihong, et al. Source identification for river sudden water contamination based on coupled probability density function method[J]. Journal of Hydraulic Engineering,2015,46(11): 1280-1289. (in Chinese))

[13] 杨海东,肖宜,王卓民,等. 突发性水污染事件溯源方法[J]. 水科学进展,2014,25(1):123-129. (YANG Haidong, XIAO Yi, WANG Zhuomin, et al. Our source identification method for sudden water pollution accidents[J]. Advances in Water Science, 2014,25(1): 123-129. (in Chinese))

[14] 沐阿华,周绍磊,于晓丽. 一种快速自适应遗传算法及其仿真研究[J]. 系统仿真学报,2004,10(1):122-125. (MU Ahua, ZHOU Shaolei, YU Xiaoli. Research on fast self-adaptive genetic algorithm and its simulation[J]. Journal of System Simulation, 2004,10(1): 122-125. (in Chinese))

[15] 余常昭. 环境流体力学导论[M]. 北京:清华大学出版社,1992.

[16] 李晓,唐洪武,王玲玲,等. 平原河网地区闸泵群联合调度水环境模拟[J]. 河海大学学报(自然科学版),2016,44(5): 393-399. (LI Xiao, TANG Hongwu, WANG Lingling, et al. Simulation of water environment under joint operation of gates and pumps in plain river network area[J]. Journal of Hohai University (Natural Sciences),2016,44(5):393-399. (in Chinese))

[17] 范玉科,侯为根,徐浩. 河道污染的一维对流-扩散方程源项识别反问题[J]. 安徽工业大学学报(自然科学版),2014,31(3):323-327. (FAN Yuke, HOU Weigen, XU Hao. River pollution source identification inverse problem of the one dimensional convection-diffusion equation[J]. Journal of Anhui University of Technology (Natural Science),2014,31(3): 323-327. (in Chinese))

[18] CHENG Weiping, JIA Yafei. Identification of contaminant point source in surface waters based on backward location probability density function method[J]. Advances in Water Resources,2010,33(4): 397-410.

[19] 杨启文,蔡亮,薛云灿. 差分进化算法综述[J]. 模式识别与人工智能,2008,21(4):506-513. (YANG Qiwen, CAI Liang, XUE Yuncan. A survey of differential evolution algorithms[J]. Pattern Recognition and Artificial Intelligence,2008,21(4): 506-513. (in Chinese))

(收稿日期:2017-07-07 编辑:高建群)



· 简讯 ·

水资源、水环境及水生态科学与工程论坛将在南京召开

为加快生态文明体制改革,推进节水型社会建设和水生态文明建设,推进水资源全面节约和循环利用,落实绿色发展理念,平衡江河湖海的保护与开发,由水资源高效利用与工程安全国家研究中心主办,河海大学、中国水利学会、中国水利企业协会等支持,南京河海科技有限公司、江苏省非常规水源开发利用工程技术中心承办的水资源、水环境及水生态科学与工程论坛将于2018年11月在南京召开。

论坛主题:水资源优化配置、水资源精细化管理、综合节水、非常规水资源开发利用、水环境治理与保护、智慧水务、河湖管理与水生态修复等。

(本刊编辑部供稿)