

Study on Prediction of Saltwater Intrusion Based on Support Vector Machine in the Pearl River Estuary

Huabin Ke¹, Bingjun Liu^{1*}, Wen Zhang¹, Zhongce Chen²

¹Department of Water Resources and Environment, School of Geography and Planning, Sun Yat-sen University, Guangzhou Guangdong

²Guangdong Water Resources and Hydropower Technology Center, Guangzhou Guangdong
Email: 149449192@qq.com, liubj@mail.sysu.edu.cn

Received: Dec. 16th, 2014; accepted: Dec. 30th, 2014; published: Jan. 16th, 2015

Copyright © 2015 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Effected by multiple factors such as upstream runoff, tide, topography, and wind, the saltwater intrusion of Pearl River Estuary displays extreme complexity. In this paper, a hybrid method based on Support Vector Machine (SVR) and Classification (SVC) is used to build a daily average salinity prediction model in the Pearl River Estuary which has considered the variation of upstream runoff and tide. The result shows that the model for saltwater intrusion prediction has high accuracy and only requires a few samples. Furthermore, the model has a good ability of self-learning and associative memory which can well simulate the complex saltwater intrusions of this estuary. The findings of this study also show that the previous average three-day runoff from upstream and the previous three-day tidal range from outlet have a good time correlation with saltwater intrusion in this estuary.

Keywords

Pearl River Estuary, Saltwater Intrusion, Support Vector Machine

基于支持向量机的珠江河口咸潮上溯预测研究

柯华斌¹, 刘丙军^{1*}, 章文¹, 陈仲策²

作者简介: 柯华斌(1988-), 男, 硕士研究生, 从事水文水资源研究方向。

*通讯作者。

¹中山大学地理科学与规划学院水资源与环境系, 广东 广州

²广东省水利水电技术中心, 广东 广州

Email: 149449192@qq.com, liubj@mail.sysu.edu.cn

收稿日期: 2014年12月16日; 录用日期: 2014年12月30日; 发布日期: 2015年1月16日

摘要

受径流、潮汐、地形、风向等多重因素影响, 珠江河口区咸潮上溯现象异常复杂。本文运用支持向量机回归(SVR)与分类(SVC)联合建模的方法, 综合考虑上游径流和河口区潮汐因素, 构建珠江河口区磨刀门水道咸潮上溯预测模型, 并采用动态反馈机制分别开展了磨刀门水道日均咸度和超标历时预测。研究表明: 基于支持向量机的咸潮上溯模型, 具有模型简单精度高, 样本容量要求少, 自学习与联想记忆能力强的特点, 能较好模拟预测多重因素影响的复杂河口区咸潮上溯现象; 磨刀门水道上游三水马口前三日流量、河口区前三日三灶站的平均潮差过程与咸潮上溯具有较好时间相关性。

关键词

珠江河口区, 咸潮上溯, 支持向量机

1. 引言

受上游径流、河口区潮汐动力、河口地形、风力风向等多重因素影响[1] [2], 河口区咸潮上溯具有高度非线性与复杂性特征[3] [4]。研究预测多重因素影响下河口区咸潮上溯, 对河口区淡水资源的开发利用, 保障供水安全具有重要的理论和实践意义。

当前, 咸潮上溯预测模型研究主要包括数学模型与物理模型两类。现行数学模型多以数值模拟为主, 难以准确描述众多因素对咸潮入侵的影响, 如一维或二维数值模型都因无法模拟垂向密度差异等因素的影响导致咸潮预测效果欠佳[5]; 理论模式中 savenije 涨憩模式虽然能快速模拟预报一个潮周期内最大咸度及咸潮最大影响范围, 但所需资料参数过多导致推广应用难度大[6]; 物理模型则以大量实地观测为主, 耗时长, 费用高, 成本大。由 Vapnik [7]提出的基于统计学习理论的支持向量机(Support Vector Machine, SVM), 以结构风险最小化原则为基础, 在处理小样本、非线性问题上具有明显的优势, 克服了人工神经网络的过学习、局部优化等问题。目前, 支持向量机方法已经广泛运用于洪水预报、径流预测、地下水预报等领域, 并逐步运用于河口区咸潮上溯预报(刘德地[8])。

本文从统计学习理论角度, 利用支持向量机对小样本、非线性问题的自组织、自学习和联想记忆等优势, 综合考虑径流、潮汐对咸潮上溯的影响, 运用支持向量机回归(SVR)和分类(SVC)联合建模的方法, 建立了咸潮上溯预测模型, 并采用动态反馈机制分别对逐日最大咸度和超标历时进行了预测。

2. 研究背景区

本文选取平岗站为研究站点, 平岗站位于磨刀门水道中游河段, 距出海口 35 km, 该站点每年的盐度序列数据代表性良好, 能较好反映每年咸潮入侵磨刀门水道的基本情况, 是珠海市重要的取水泵站。根据珠江三角洲的水力条件, 选用马口水文站与三水水文站的日均合流量(简称马 + 三流量)作为上游来水资料, 马口站及三水站具备长序列流量资料, 其径流量变化特征具有良好的代表性。下游潮汐资料选用三灶站, 该站距离磨刀门 16 km 左右, 其潮位特征同时受外海潮汐、上游径流来水来沙以及河口区地

形条件影响，是珠江河口区最重要验潮站之一。磨刀门水道及各测站位置示意图见图 1。

3. 咸潮上溯预测模型

3.1. 支持向量机基本原理[7]

支持向量机是建立在统计学习理论的 VC 维理论和结构风险最小化原则上的，根据有限的样本信息在模型复杂性与学习能力之间寻求最佳折衷，以求获得最佳推广能力的新型学习方法。它具有回归拟合与分类识别功能，在解决小样本、非线性及高维问题中有很强的优势。

1) 支持向量机回归(SVR)

支持向量机回归问题是利用非线性变换将样本映射到高维特征空间(Hilbert 空间)，从而将输入空间中的非线性函数估计问题转化为高维特征空间中的线性函数估计问题。同时将其最小化风险函数等价如下优化问题：

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (1)$$

$$\text{s.t.} \begin{cases} \omega^T \phi(x_i) + b - y_i \leq \varepsilon + \xi_i \\ y_i - \omega^T \phi(x_i) - b \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, l \end{cases} \quad (2)$$

式中， $\phi(x_i)$ 为输入的非线性映射集合， ω 为权向量， C 为惩罚因子， b 为偏置量， ε 为不敏感参数， ξ_i, ξ_i^* 为松弛变量。利用对偶原理、拉格朗日乘子法和核函数技术，上述优化问题转化为对偶问题并求解，最终可得到回归函数如下，其中 $K(x_i, x)$ 为一个满足 Mercer 条件的核函数。

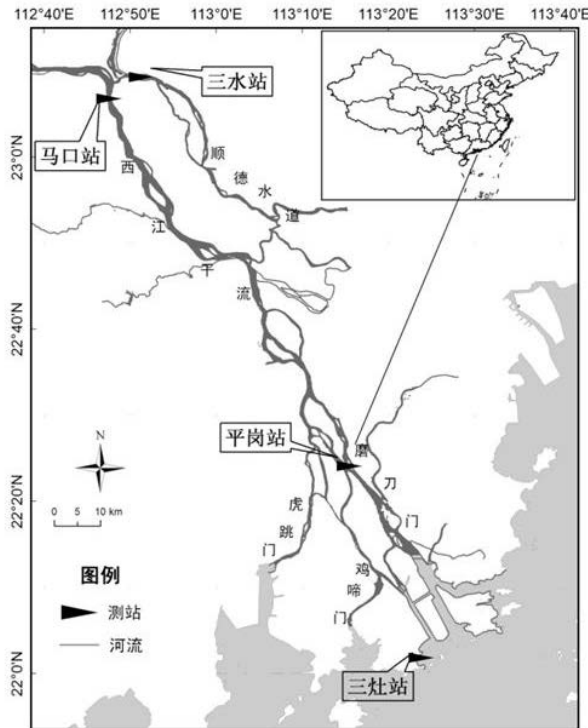


Figure 1. Location of research region

图 1. 研究区域位置图

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (3)$$

2) 支持向量机分类(SVC)

支持向量机分类问题原理与 SVR 类似，将输入空间通过非线性变换映射到高维特征空间，构造线性的最优分类超平面。其最小化风险函数可等价如下优化问题：

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i \quad (4)$$

$$\text{s.t.} \begin{cases} y_i (\omega^T \phi(x_i) + b) \geq 1 - \xi_i \\ \xi_i \geq 0, i = 1, \dots, l \end{cases} \quad (5)$$

式中， $\phi(x_i)$ 为输入的非线性映射集合， ω 为权向量， C 为惩罚因子， b 为偏置量， ξ_i 为松弛变量。利用对偶原理、拉格朗日乘子法和核函数技术，上述优化问题转化为对偶问题并求解，最终得到分类函数为：

$$f(x) = \text{sgn} \left(\sum_{i=1}^l y_i \alpha_i K(x_i, x) + b \right) \quad (6)$$

3.2. 咸潮上溯预测模型构建

1) 基于 SVR 的日均咸度预测模型

珠江口咸潮上溯受上游径流和下游潮汐作用明显[9]，在咸度预测模型输入端需考虑上游径流因素(上游马 + 三流量)与下游潮汐作用(三灶站日潮差过程)的影响；另一方面，路剑飞等[10]发现咸度序列具有一定自相关性，滞后步长在 2 d 以内自相关系数超过 80%。因此将前两日的历史咸度资料也作为模型输入条件。输出端为预测当天的平岗站日均咸度。日均咸度预测模型如下：

$$y_t = f(x, z, y_{t-2}, y_{t-1}) \quad (7)$$

式中， x 为马 + 三流量， z 为三灶站日潮差， y_{t-1} 、 y_{t-2} 为第 $t-1$ 、 $t-2$ 天平岗站的日均咸度， y_t 为预测当天日均咸度。

为消除变量量纲影响，减少训练过程中数值运算的复杂程度，采用比例压缩法对样本数据归一化处理；选取高斯径向基(RBF)函数作为核函数[11]；利用 Grid-Search 方法搜索模型的主要参数惩罚因子 C 与核函数参数 g ，搜索区间为 $[2^{-8}, 2^8]$ ，步长为 0.5；运用 K 折交叉验证法(k-fold cross-validation) [12]检验参数。

2) 基于 SVC 的超标历时预测模型

运用支持向量机分类(SVC)的识别功能，对超标历时进行识别分类，从而预测超标历时过程，分类标准见表 1。模型输入端考虑上游马 + 三流量、下游三灶站日潮差过程、平岗站的日均咸度等历史资料；输出端为预测当天平岗站的日超标历时类别。超标历时预测模型如下：

$$C_t = f(x, z, y_{t-2}, y_{t-1}, y_t) \quad (10)$$

Table 1. The classification criteria of hours for daily exceeding standards salinity
表 1. 日超标历时分类标准

超标历时(h)	0~4	5~9	10~14	15~19	20~24
类别(级别)	1	2	3	4	5

式中, x 为马 + 三流量, z 为三灶站日潮差, y_{t-1} 、 y_{t-2} 为第 $t-1$ 、 $t-2$ 天平岗站的日均咸度, y_t 为预测当天日均咸度, 这里引入日均咸度预测模型的预测结果; C_t 为平岗站预测当天的超标历时分类。

3.3. 模型检验指标

为评价模型的拟合效果, 分别选用效果系数、一致性指标、分类准确率等三个拟合度量指标表征。

1) “效果系数”(coefficient of efficiency), 用 E 表示。效果系数 E 愈接近 1, 则说明预测值愈接近实测值, 但效果系数 E 对系列中的极值是敏感的, 因此需要结合一致性指标 D 判断来消除此影响。

$$E = 1.0 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (11)$$

2) “一致性指标”(index of agreement), 用 D 表示。一致性指标 D 的取值在 0~1 之间, 取值越大, 说明模型预测值变化与实测值越一致。

$$D = 1.0 - \frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (12)$$

3) 分类准确率, 用 T 表示。是用来衡量模型分类结果准确性的指标。

$$T = \text{length}(|P'_i = O'_i|) / N \times 100\% \quad (13)$$

式中, O 为实测值, P 为预测值, \bar{O} 为实测平均值, O' 为实测值类别, P' 为预测值类别, N 为样本数。

4. 结果分析

选取 2003 年 12 月 1 日至 2004 年 2 月 29 日的实测资料(上游马 + 三流量、三灶站日潮差与平岗站日均咸度同步资料)为计算样本。其中, 选用 2003 年 12 月 1 日至 2004 年 1 月 31 日的数据作为训练集, 选取 2004 年 2 月 1 日至 2 月 29 日共 29 天的数据作为模型检验样本。

4.1. 基于 SVR 的日均咸度预测模型结果分析

比较不同径流、潮汐序列组合输入的模型模拟结果(见表 2), 发现当输入条件选择前三日马 + 三流量(x_{t-3})与前三日三灶站日潮差的平均值(z_m)时, 模型拟合效果最好, 效果系数 E 达到 0.9081, 一致性指标 D 达到 0.9749。因此, 确定该组合为最佳输入条件; 经过参数优选, 确定模型参数 $C = 256$, $g = 0.5$,

Table 2. Simulation of different model input conditions

表 2. 不同径流与潮差组合输入的模型模拟结果对比

输入变量	x_{t-3}		x_{t-2}		x_{t-1}		x_m	
	E	D	E	D	E	D	E	D
z_{t-3}	0.8739	0.9656	0.7367	0.9194	0.7631	0.9292	0.7247	0.914
z_{t-2}	0.8217	0.9448	0.8449	0.9556	0.8767	0.9664	0.8399	0.9541
z_{t-1}	0.7921	0.9369	0.8156	0.9452	0.7887	0.9375	0.8176	0.9458
z_m	0.9081	0.9749	0.8362	0.9515	0.8804	0.967	0.8145	0.9444

注: x_{t-3} 、 x_{t-2} 、 x_{t-1} 分别为前三天的马+三流量, x_m 为 x_{t-3} 、 x_{t-2} 、 x_{t-1} 的平均值。 z_{t-3} 、 z_{t-2} 、 z_{t-1} 分别为前三天的三灶站潮差, z_m 为 z_{t-3} 、 z_{t-2} 、 z_{t-1} 的平均值。

不敏感参数 ε 取 0.001。具体参数优选和模拟结果见图 2。

为分析模型的预测精度，分别对比了普通 SVR 模型(未采用动态反馈机制)、常用 BP 神经网络模型和动态反馈 SVR 模型的预测结果，结果见图 3。其中，动态反馈机制，是将模型执行过一天的预测，作为最新的当天实测数据加入训练集并重新训练模型，通过动态反馈机制使模型在使用过程中不断提炼。分析图 4，动态反馈 SVR 模型的预测效果最好，测试集效果系数 E 达到 0.9779，一致性指标 D 达到 0.9941；普通 SVR 模型测试集效果系数 E 只有 0.8167，一致性指标 D 只有 0.9391；由于样本数目较少，导致 BP 神经网络模型的训练神经网络并不成熟，泛化能力较差，其测试集效果系数 E 只有 0.7706，一致性指标 D 也只有 0.9118。由此可见，支持向量机在小样本的学习问题更为出色，动态反馈机制也能有效地提高模型预测精度。

4.2. 基于 SVC 的超标历时预测模型结果分析

比较不同径流、潮汐组合输入的超标历时预测模型模拟结果，见表 3。当输入选择前三日马 + 三流

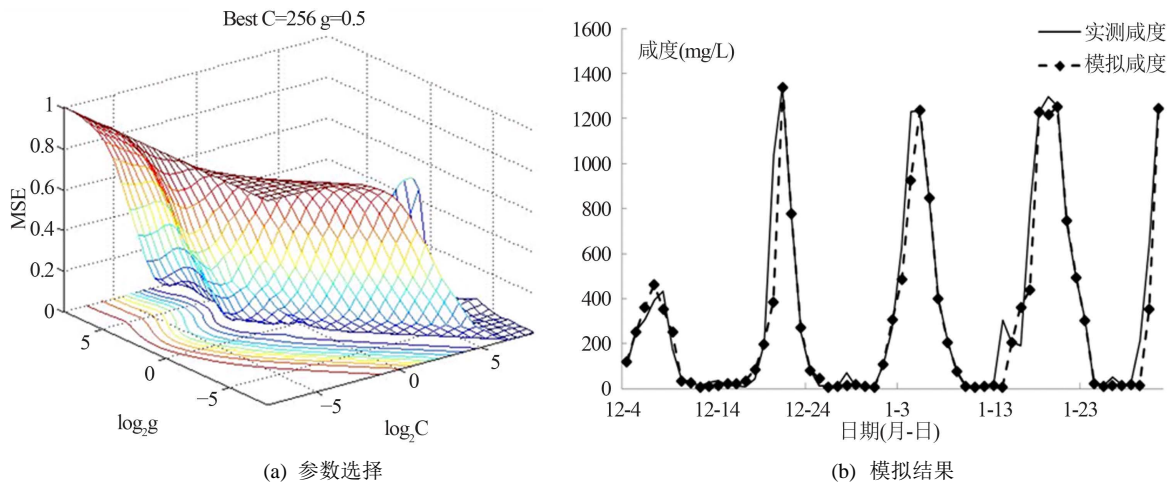


Figure 2. The parameter selection and simulation results for average daily salinity prediction model
图 2. 日均咸度预测模型参数优选与模拟结果

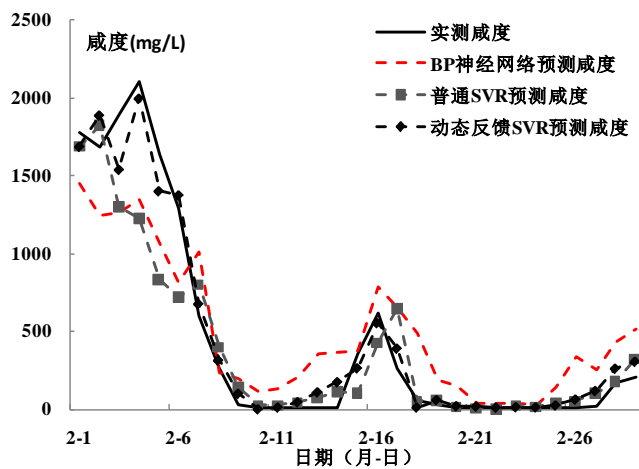


Figure 3. The comparison of different model prediction for average daily salinity
图 3. 日均盐度模型预测结果对比图

量(x_{t-3})与前三日三灶站日潮差的平均值(z_m)时,模型模拟准确率最高,分类准确率 T 达到 100%。经过参数优选,确定模型参数 $C = 1$, $g = 5.6569$ 。具体参数优选和模拟结果见图 4。

加入动态反馈机制,分析普通 SVC 模型、常用 BP 神经网络模型和动态反馈 SVC 模型的预测效果(图 5)。

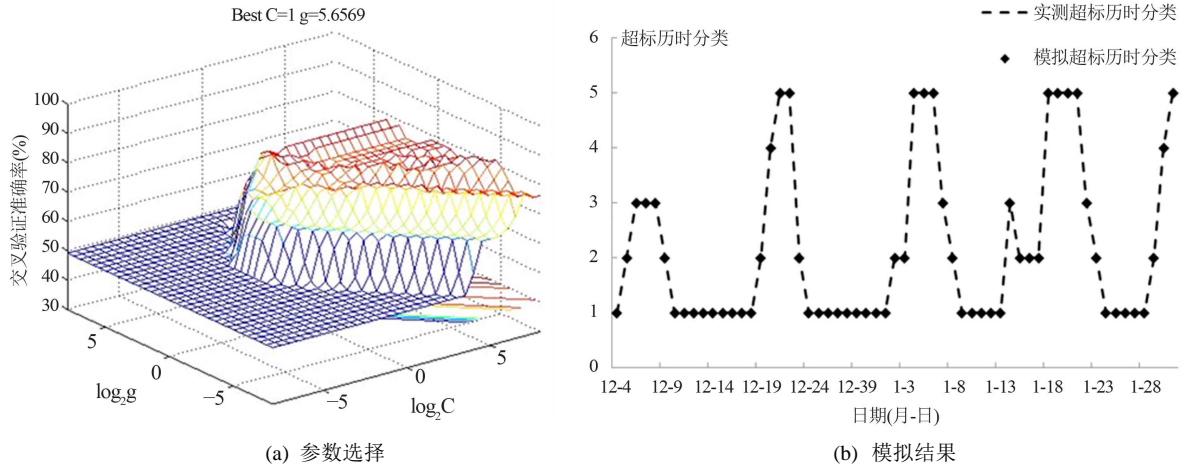


Figure 4. The parameter selection and simulation results for the prediction model of exceeding standard salinity
图 4. 日均咸度超标历时模型参数与模拟结果

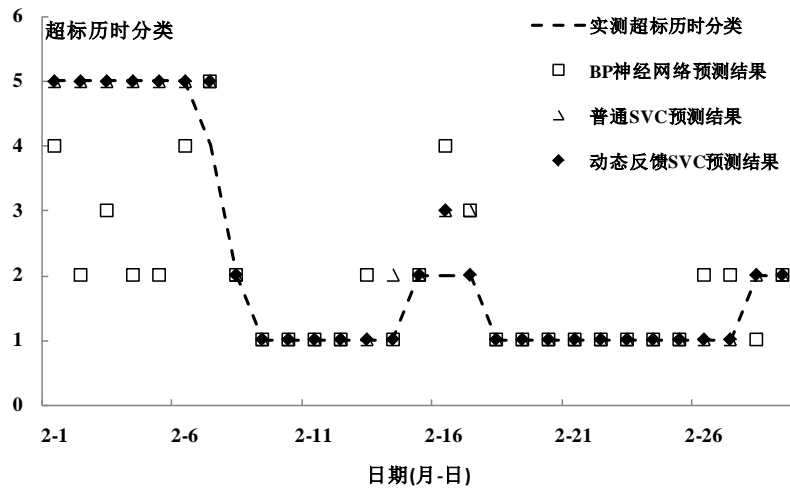


Figure 5. The comparison of different model prediction for the prediction model of exceeding standard salinity
图 5. 日均盐度超标历时模型预测结果对比图

Table 3. The classification accuracy of different model input conditions
表 3. 不同径流与潮差组合输入的超标历时分类模拟结果(T) (单位: %)

输入变量	x_{t-3}	x_{t-2}	x_{t-1}	x_m
z_{t-3}	98.31%	98.31%	98.31%	98.31%
z_{t-2}	98.31%	98.31%	98.31%	98.31%
z_{t-1}	98.31%	98.31%	98.31%	98.31%
z_m	100%	98.31%	98.31%	98.31%

可以得出, 动态反馈 SVC 模型的测试集预测分类准确率最理想, 达到 93.10%, 较普通 SVC 模型分类准确率提高了 6.90%; BP 神经网络模型分类预测效果明显欠佳, 分类准确率只有 55.2%。表明动态反馈 SVC 模型能更好地预测超标历时分类。

5. 结论

受径流、潮汐、地形、风向等多重因素影响, 珠江河口区咸潮上溯现象异常复杂。本文尝试运用支持向量机回归(SVR)与分类(SVC)联合建模方法, 综合考虑上游径流和河口区潮汐因素, 建立珠江河口区磨刀门水道咸潮上溯预测模型, 得出以下结论:

1) 支持向量机方法具有很强的自组织、自学习与联想记忆能力, 能较好应用于咸潮上溯这种资料较为缺乏的小样本、非线性问题。本文构建的基于支持向量机的咸潮上溯预测模型, 采用动态反馈机制, 能较好模拟磨刀门水道平岗站日均咸度和超标历时, 模型预测效果明显优于人工神经网络。

2) 通过比较模型不同径流、潮汐组合输入的模型计算结果, 确定前三日马+三流量序列和前三日三灶站平均潮差序列为最佳输入条件。

基金项目

广东省水利科技创新项目(2013-13); 国家自然科学基金项目(50909106, 51009156); 广东省科技计划项目(2011B0308 00008)。

参考文献 (References)

- [1] ZHANG X.-F., DENG J.-Q. Affecting factors of salinity intrusion in Pearl River Estuary and sustainable utilization of water resources in Pearl River Delta. Sustainability in food and water. Netherlands: Springer, 2010: 11-17.
- [2] 宋志尧, 茅丽华. 长江口盐水入侵研究[J]. 水资源保护, 2002, 3, 27-30.
SONG Zhiyao, MAO Lihua. Saltwater intrusion at Yangtze River Estuary. Water Resources Protection, 2002, 3: 27-30. (in Chinese)
- [3] 章文, 刘丙军, 辛彦博, 等. 珠江河口区盐度变化周期特征分析——以磨刀门水道为例[J]. 热带地理, 2013, 33(1): 28-33.
ZHANG Wen, LIU Bingjun, XIN Yanbo, et al. Analysis of the characters of saltwater intrusion in Pearl River Estuary based on salinity data of Modaomen Waterway. Tropical Geography, 2013, 33(1): 28-33. (in Chinese)
- [4] 闻平, 陈晓宏, 刘斌, 等. 磨刀门水道咸潮入侵及其变异分析[J]. 水文, 2007, 27(3): 65-67.
WEN Ping, Chen Xiaohong, LIU Bin, et al. Analysis of tidal saltwater intrusion and its variation in Modaomen Channel. Journal of China Hydrology, 2007, 27(3): 65-67. (in Chinese)
- [5] 欧素英, 杨青书, 雷亚平. 咸潮入侵理论预报模式的分析及在西江三角洲的应用[J]. 热带海洋学报, 2010, 29(1): 32-41.
OU Suying, YANG Qingshu and LEI Yaping. Analysis of a model for saline water intrusion and its application in the West River Delta. Journal of Tropical Oceanography, 2010, 29(1): 32-41. (in Chinese)
- [6] SAVENIJE, H. H. G. Salinity and tides in alluvial estuaries. Amsterdam: Elsevier, 2005.
- [7] VAPNIK, V. The nature of statistical learning theory. New York: Springer-Verlag, 1995.
- [8] 刘德地, 陈晓宏. 基于偏最小二乘回归与支持向量机耦合的咸潮预报模型[J]. 中山大学学报(自然科学版), 2007, 46(4): 89-92.
LIU Dedi, CHEN Xiaohong. Model for prediction of saltwater intrusions based on coupling of support vector machine and partial least square method. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2007, 46(4): 89-92. (in Chinese)
- [9] GONG, W., SHEN, J. The response of salt intrusion to changes in river discharge and tidal mixing during the dry season in the Modaomen Estuary, China. Continental Shelf Research, 2011, 31(7): 769-788.
- [10] 路剑飞, 陈子燊. 珠江口磨刀门水道盐度多步预测研究[J]. 水文, 2010, 30(5): 69-74.
LU Jianfei, CHEN Zishen. Salinity prediction at Modaomen Waterway in estuary of Pearl River. Journal of China Hydrology, 2010, 30(5): 69-74. (in Chinese)
- [11] HSU, C. W., CHANG, C. C. and LIN, C. J. A practical guide to support vector classification. Technical Report, De-

partment of Computer Science and Information Engineering, University of National Taiwan, Taipei, 2003: 1-12.

- [12] KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. IJCAI, 1995, 14(2): 1137-1145.