

论著·公共卫生

重庆市 2009—2016 年细菌性痢疾空间流行病学特征及基于气象要素的预测模型研究

刘 勋¹, 孟秋雨¹, 谢佳伽¹, 肖达勇², 王 怡¹, 邓 丹¹

1. 重庆医科大学公共卫生与管理学院, 医学与社会发展研究中心, 健康领域社会风险预测治理协同创新中心, 重庆 400016; 2. 重庆市疾病预防控制中心地方病与寄生虫病预防控制所, 重庆 400042

[摘要] **目的**·分析重庆市细菌性痢疾的空间流行病学特征及其与气象要素的相关性, 并构建其发病率预测模型, 为重庆市细菌性痢疾疫情的防控提供科学依据。**方法**·收集 2009—2016 年重庆市细菌性痢疾及气象要素数据, 并进行描述性流行病学分析, 采用时空扫描统计量进行细菌性痢疾时空聚集性分析, 运用 DCCA 系数法量化细菌性痢疾发病率与气象要素的相关性, 运用 Boruta 算法结合粒子群优化算法 (particle swarm optimization, PSO) 及支持向量机回归模型 (support vector machine for regression, SVR) 构建细菌性痢疾发病率预测模型。**结果**·① 2009—2016 年重庆市细菌性痢疾年均报告发病率为 29.394/100 000, 0~5 岁年龄组发病率 (295.892/100 000) 最高, 散居儿童占比 (50.335%) 最大, 5 月—10 月为其季节性发病高峰; 细菌性痢疾呈现显著的时空聚集性, 一类聚集区主要集中在重庆市主城区, 二类聚集区主要集中在重庆市东北地区; 6 月—10 月为其主要的聚集时间。② 与人群细菌性痢疾发病率具有很强相关性的气象要素分别为月平均气压 ($\rho_{DCCA}=-0.918$)、月平均最高气温 ($\rho_{DCCA}=0.875$) 及月平均气温 ($\rho_{DCCA}=0.870$)。③ 基于气象要素构建的 PSO_SVR 模型均方误差 (mean squared error, MSE)、平均绝对百分比误差 (mean absolute percentage error, MAPE)、平方相关系数 (square correlation coefficient, R^2) 分别为 0.055、0.101 及 0.909。**结论**·重庆市主城区及渝东北地区应作为细菌性痢疾的重点防控区域, 同时相关卫生部门应结合气象要素与细菌性痢疾发病率的密切相关性及其季节性高发特点, 对 0~5 岁儿童、散居儿童、农民等人群采取针对性的应对措施以控制细菌性痢疾传播与流行。基于气象要素建立的 PSO_SVR 模型预测性能良好, 可为细菌性痢疾的防控提供有力的理论支撑。

[关键词] 细菌性痢疾; 空间流行病学; DCCA 系数法; Boruta 算法; 粒子群优化算法; 支持向量机回归模型

[DOI] 10.3969/j.issn.1674-8115.2019.02.015 **[中图分类号]** R183.4 **[文献标志码]** A

Spatial epidemiological characteristics and prediction models of bacterial dysentery in Chongqing from 2009 to 2016 based on meteorological elements

LIU Xun¹, MENG Qiu-yu¹, XIE Jia-jia¹, XIAO DA-yong², WANG Yi¹, DENG Dan¹

1. School of Public Health and Management, Research Center for Medicine and Social Development, Innovation Center for Social Risk Governance in Health, Chongqing Medical University, Chongqing 400016, China; 2. Institute for Prevention and Control of Endemic and Parasitic Diseases, Chongqing Center for Disease Control and Prevention, Chongqing 400042, China

[Abstract] **Objective**·To analyze the spatial epidemiological characteristics of bacillary dysentery and its correlation with meteorological elements in Chongqing, and to construct its incidence prediction model, thus providing scientific basis for the prevention and control of bacterial dysentery. **Methods**·The data of bacterial dysentery cases and meteorological factors from 2009 to 2016 in Chongqing was collected in this study. Descriptive methods were employed to investigate the epidemiological distribution of bacillary dysentery. Spatiotemporal scanning statistics was used to analyze spatiotemporal characteristics of bacillary dysentery. DCCA coefficient method was used to quantify the correlation between the incidence of bacillary dysentery and meteorological elements. Both Boruta algorithm and particle swarm optimization algorithm (PSO) combined with support vector machine for regression model (SVR) were used to establish the prediction model for the incidence of bacterial dysentery. **Results**·① The mean annual reported incidence of bacillary dysentery in Chongqing from 2009 to 2016 was 29.394/100 000. Children <5 years old had the highest incidence (295.892/100 000) among all age categories and scattered children had the highest proportion (50.335%) among all occupation categories. The seasonal incidence peak was from May to October. Bacterial dysentery showed a significant spatial-temporal aggregation that the most likely clusters for disease was found mainly in the main urban areas and main gathering time was from June to October. ② The most important meteorological elements associated with the incidence of bacterial dysentery were monthly mean atmospheric pressure ($\rho_{DCCA}=-0.918$), monthly mean maximum temperature ($\rho_{DCCA}=0.875$) and monthly mean temperature ($\rho_{DCCA}=0.870$). ③ The mean squared error (MSE), mean absolute percentage error (MAPE) and square correlation coefficient (R^2) of PSO_SVR model constructed based on meteorological elements were 0.055, 0.101 and 0.909, respectively. **Conclusion**·The main urban areas of Chongqing and the northeast of Chongqing should be regarded as the key areas for the prevention and control of bacillary dysentery. At the same time, according to the characteristics of bacillary dysentery, relevant health departments should take targeted measures to control the spread and prevalence of bacillary dysentery among children <5 years old, scattered children and farmers. The PSO_SVR model constructed based on meteorological elements has good predictive performance and can provide scientific theoretical support for the prevention and control of bacterial dysentery.

[Key words] bacterial dysentery; spatial epidemiology; DCCA coefficient method; Boruta algorithm; particle swarm optimization algorithm; support vector machine regression

[基金项目] 重庆市基础研究与前沿探索项目 (cstc2018jcyjA0135); 重庆市卫生和计划生育委员会 2015 年医学科研计划项目 (2015MSXM094) (Basic Research and Frontier Exploration Project in Chongqing, cstc2018jcyjA0135; Medical Research Project of Chongqing Health and Family Planning Commission in 2015, 2015MSXM094)。

[作者简介] 刘 勋 (1991—), 男, 硕士生; 电子信箱: 564658790@qq.com。

[通信作者] 邓 丹, 电子信箱: 100079@cqmu.edu.cn。

细菌性痢疾 (bacillary dysentery), 简称菌痢, 是一种由志贺菌属引起的严重肠道传染病。其发病率一直位居我国法定传染病前列, 具有流行范围广、传播速度快、对健康危害大等特点, 已给社会带来了严重的疾病负担。重庆市地处中国西南地区, 属亚热带季风性湿润气候, 境内地貌以丘陵、山地为主, 河流众多, 雨水充沛, 利于志贺菌属的繁殖与传播。研究显示, 近年来重庆市个别地区菌痢发病率高达 103.21/100 000^[1], 而全国同期菌痢发病率为 15.29/100 000^[2]。由此可见, 目前重庆市菌痢疫情依然十分严峻, 加强菌痢的防治工作刻不容缓。鉴于气候变化对传染病发生、传播过程的重要影响, 已有诸多学者^[3-5]采用了如 Pearson 相关系数、多因素线性回归等传统统计学方法, 探讨了气象要素对传染病发病率的影响与相关性; 然而, 上述方法仅适于具有平稳特征的研究数据, 而当研究数据具有非平稳特性时, 相关方法所得到的结论不可信^[6]。同时, 传染病的发生、发展通常具有空间分布规律, 以往研究大多集中在流行病学特征描述及病原学分析, 而忽视了传染病的空间分布特征。

因此, 本研究拟采用 DCCA 系数法量化气象要素与菌痢发病率的相关性, 同时采用空间分析技术, 分析重庆市菌痢的流行病学特征及时空聚集情况, 在前两部分研究基础之上运用机器学习方法构建菌痢发病率预测模型。通过本研究, 以期重庆市菌痢防控、资源合理配置的及时性与有效性提供有力的理论支撑。

1 资料与方法

1.1 资料来源

2009—2016 年重庆市菌痢监测数据来源于中国疾病预防控制中心信息系统, 其内容包括报告病例的性别、年龄、报告地点、发病时间、职业等信息; 同期气象数据来自重庆市气象局, 其内容包括日平均温度、日平均湿度、日平均气压等共 10 个气象要素; 重庆市地理数据 (区县级别) 来自国家基础地理信息中心网站。

1.2 研究方法

1.2.1 空间流行病学特征分析 采用描述性流行病学方法分析菌痢三间分布特征; 以区县为最小空间尺度, 采用时空扫描统计量对 2009—2016 年重庆市菌痢数据按月进行时空聚集性分析。分析时, 设定扫描步长为 1 个月, 最大扫描半径为风险人口数的 20%, 选用圆柱形窗口对研究区域进行扫描; 对于每个扫描窗口, 根据实际发病数和期望发病数计算其最大似然比 (loglikelihood ratio, LLR)

和相对危险度 (relative risk, RR), 以此判断扫描窗口内疾病是否存在时空聚集及其聚集程度。设定检验水准取 $\alpha=0.05$, 蒙特卡罗随机试验次数为 999。当 LLR 值具有统计学意义时, 表明扫描窗口内菌痢发病存在时空聚集^[7]。

1.2.2 相关性分析 采用 DCCA 系数法量化气候要素与菌痢发病率的相关性水平。DCCA 系数法是一种可用于度量 2 个非平稳序列的相关性的研究方法^[8], 其具体算法如下。

给定时间序列 $X(i)$ 和 $Y(i)$, 其中 $i=1, 2, \dots, N$, 构造新序列 $x(k) = \sum_{i=1}^k (x_i - \bar{x})$, $y(k) = \sum_{i=1}^k (y_i - \bar{y})$, 其中 $k=1, 2, \dots, N$, \bar{x} 和 \bar{y} 为原始序列 $X(i)$ 和 $Y(i)$ 的平均值。将新的时间序列 $x(k)$, $y(k)$ 划分为 $N-s$ 个重叠的子区间, 每个子区间包含 $s+1$ 个值; 然后, 利用最小二乘法估计法对每一个子序列的局部趋势进行拟合, 得到 $\tilde{x}_{k,i}$, $\tilde{y}_{k,i}$; 接着计算每一个子序列的局部降趋势协方差 $f_{DCCA}^2(s, i) = 1/(s+1) \sum_{k=i}^{i+s} (x_k - \tilde{x}_{k,i})(y_k - \tilde{y}_{k,i})$, 再将 $N-s$ 个子区间的局部降趋势协方差取均值得到 $f_{DCCA}^2(s) = (N-s)^{-1} \sum_{i=1}^{N-s} f_{DCCA}^2(s, i)$ 。最后, 定义 DCCA 系数 ρ_{DCCA} 为:

$$\rho_{DCCA}(s) = \frac{f_{DCCA}^2(s)}{f_{DFA}(x_i) f_{DFA}(y_i)}$$

其中 $f_{DFA}(x_i)$ 、 $f_{DFA}(y_i)$ 为序列 $X(i)$ 和 $Y(i)$ 的降趋势方差函数, 其计算过程与 $f_{DCCA}^2(s)$ 类似; 例如序列 $X(i)$ 的降趋势方差函数 $f_{DFA}(s) = \{ \frac{1}{N-s} \sum_{i=1}^{N-s} f_{DFA}^2(s, i) \}^{1/2}$, 式中 $f_{DFA}^2(s, i) = 1/(s+1) \sum_{k=i}^{i+s} (x_k - \tilde{x}_{k,i})^2$ 。 ρ_{DCCA} 为无量纲的量, 且 $-1 \leq \rho_{DCCA} \leq 1$, $\rho_{DCCA}=1$, $\rho_{DCCA}=-1$ 和 $\rho_{DCCA}=0$ 分别表示 2 个序列完全正相关、负相关和无相关性。

1.2.3 构建发病率预测模型 将菌痢月发病率作为模型预测变量 (因变量), 将月发病率的滞后 1 个月菌痢发病人数及气象要素作为模型自变量。选择 2009 年 1 月—2015 年 4 月共 76 个时间点的数据作为训练样本, 将 2015 年 5 月—2016 年 12 月共 20 个时间点的数据作为测试样本。选择 Boruta 算法^[9]进行特征选择, 然后根据特征选择结果, 采用基于粒子群优化算法 (particle swarm optimization, PSO) 的支持向量机 (support vector machine, SVM) 模型构建发病率预测模型。

Boruta 算法是基于随机森林算法构建的包装型特征选择算法, 可以选择出所有与因变量相关的特征集合, 而不是针对特定模型选择出可以使得模型代价函数最小的特征集合^[9]。这可以帮助研究者更全面地理解因变量的影响因素, 从而更好、更高效地进行特征选择, 提高机器学习的速率与准确性。SVM 模型是基于统计学习理论所提出的、以结构风险最小化原则为基础的机器学习方法, 而支持向量回归 (support vector machine for regression, SVR) 模型是 SVM 在回归领域的应用。对于非线性样本集

$\{(x_1, y_1), \cdots (x_m, y_m)\}$, $i=1, 2, \cdots, m$, 其中 $x_i \in R^d, y_i \in R$, SVR 首先采用核函数将低维空间线性不可分的输入变量映射到高维空间实现线性可分, 然后在样本集高维特征空间内通过使用损失函数以及引入松弛因子 ξ_i 、 ξ_i^* 和惩罚系数 C , 构造一个线性回归的凸二次规划问题, 求解该问题最终得到结构风险最小化原则下的回归函数:

$$f(x)=\sum_{i=1}^m(a_i-a_i^*)K(x_i,x)+b$$

式中 a_i 、 a_i^* 为拉格朗日乘子, b 代表阈值, $K(x_i, x)$ 为核函数。

PSO 算法是研究者受鸟群捕食行为的启发而提出的一种群体智能优化算法^[10], 具有全局优化能力强、易于实现与推广等特点, 目前已被应用到神经网络优化、函数优化、组合优化等领域^[11]。在求解优化问题时, 算法中每个粒子都代表问题的一个潜在解, 采用适应度值衡量粒子的优劣。粒子在解空间中运动, 每个粒子根据其自身历史最优适应度值所对应的位置 P_i (个体极值) 和整个粒子群的历史最优适应度值所对应的位置 P_g (群体极值) 更新自己的速度和位置, 最终通过迭代方式搜索得到最优解, 速度和位置更新公式如下:

$$V_{id}^{k+1}=wV_{id}^k+c_1r_1(P_{id}^k-X_{id}^k)+c_2r_2(P_{gd}^k-X_{id}^k)$$

$$X_{id}^{k+1}=X_{id}^k+V_{id}^{k+1}$$

式中, w 为惯性权重; $d=1, 2, \cdots, n$; $i=1, 2, \cdots, n$; k 为当前迭代次数; V_{id} 为粒子速度; X_{id} 为粒子位置; c_1 、 c_2 为加速度因子; r_1 、 r_2 为分布于 $[0, 1]$ 的随机数。

1.3 统计学分析

采用 SaTScan 9.4.2 进行菌痢空间聚集性分析; 运

用 R-3.5.1 中的 Boruta 包实现 Boruta 特征选择算法; 菌痢流行病学分析、相关性分析、预测模型的构建则是基于 MATLAB 2018a 编程完成, 选取平均绝对百分比误差 (mean absolute percentage error, $MAPE$)、均方误差 (mean squared error, MSE) 与平方相关系数 (square correlation coefficient, R^2) 作为模型预测精度评价指标。当模型 $MAPE$ 、 MSE 最小, R^2 最大时, 可认为该模型为最优预测模型。

2 结果

2.1 菌痢流行病学特征

2009—2016 年重庆市累计报告菌痢例数为 68 855, 年均发病率为 29.394/100 000, 最高发病率为 36.312/100 000, 最低发病率为 23.613/100 000, 发病率整体呈下降趋势 (图 1)。重庆市菌痢发病率前三顺位年龄组人群分别为 0 ~ 5 岁、5 ~ 10 岁及 65 岁以上人群组, 其发病率分别为 295.892/100 000、23.843/100 000 及 22.524/100 000, 发病人数分别占总发病人数的 52.879%、4.424% 及 8.925%。男性发病率高于女性, 男女平均发病率分别为 31.002/100 000 与 27.724/100 000, 发病率之比为 1.130:1, 但 10 ~ 65 岁人群女性发病率高于男性发病率。职业分布方面, 散居儿童占比最大 (34 658 例, 50.335%), 其次为农民 (8 031 例, 11.664%) 和学生 (5 592 例, 8.121%)。时间分布方面, 菌痢疫情呈现明显的季节性, 3 月份发病数开始增多; 5 月—10 月为发病高峰, 发病人数为 45 131, 占全年发病人数的 65.545%。地区分布方面, 菌痢发病高发区域主要集中在重庆市主城区及渝东北地区 (表 1)。

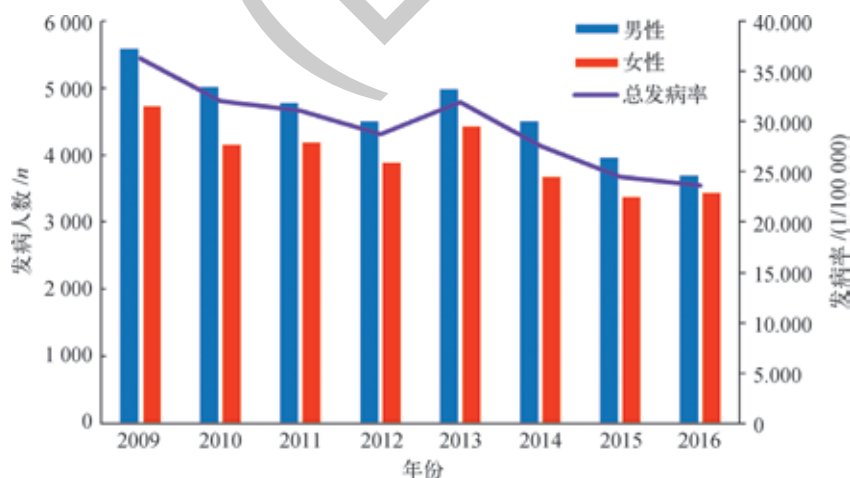


图 1 2009—2016 年重庆市菌痢总发病率及性别分布

Fig 1 Total incidences and gender distribution of bacillary dysentery in Chongqing from 2009 to 2016

表 1 2009—2016 年重庆各区县菌痢年均发病率顺位表
Tab 1 Ranking of annual incidence of bacillary dysentery in Chongqing District, 2009—2016

发病率顺位	地区	发病人数 /n	发病率 / (1/100 000)
1	九龙坡区	7 129	103.135
2	城口县	4 441	89.631
3	沙坪坝区	5 762	87.554
4	南岸区	3 007	85.540
5	渝中区	843	72.385
6	江北区	3 158	62.052
7	渝北区	5 326	60.807
8	长寿区	939	50.628
9	大渡口区	2 253	49.049
10	梁平县	1 522	32.109

2.2 时空聚集性分析

2009—2016 年重庆市菌痢发病一类聚集区域为相同的 6 个主城区, 即渝中区、北碚区、大渡口区、南岸区、江北区及九龙坡区。一类聚集区中心为 (29.55°N, 106.54°E), 聚集半径为 21.76 km, 扫描窗口内的发病率为 130.4/100 000 ~ 82.2/100 000, *RR* 值范围为 5.09 ~ 4.26; 2009—2016 年重庆市菌痢发病二类聚集区域有一定的波

动, 主要集中在渝东北地区中的部分区县, 如垫江县、丰都县、梁平县、城口县等, 除城口县外的二类聚集区域 *RR* 值范围为 1.43 ~ 3.33。需特别注意的是 2014 年、2015 年城口县 *RR* 值为 15.9 与 12.25, 明显高于其他地区。菌痢聚集时间主要为 5 月—10 月份。

2.3 相关性分析

首先采用 Podobnik 检验^[12] 对 2009—2016 年重庆市菌痢月发病率与气象要素之间的相关性进行定性检验; 结果显示, 当自由度 *m* 取值为 1 ~ 12 时, 所有气象要素 Podobnik 检验统计量 $Q_{CC}(m)$ 最小值为 $Q_{CC}(1)=21.016$, 而 $\chi^2_{0.05}(m)$ 最大值为 $\chi^2_{0.05}(12)=21.026$, 且 $Q_{CC}(2 \sim 12)$ 均大于 $\chi^2_{0.05}(12)$ 。因此, 可以在 95% 的显著性水平下拒绝 2 个非平稳时间序列之间不存在互相关性的原假设, 可认为菌痢周发病率与本研究纳入气象要素均存在相关性。进一步采用 DCCA 系数法对菌痢发病率与气象要素之间的相关性进行量化分析; 结果显示月平均最低气温、月平均水汽压、月平均气温、月平均最高气温、月平均气压与菌痢月发病率存在较强的相关性 ($|\rho_{DCCA}| \gg 0.8$), 同时月平均湿度、月平均气压与菌痢月发病率存在负相关性, 其他气象要素与菌痢发病率存在正相关性 (表 2)。

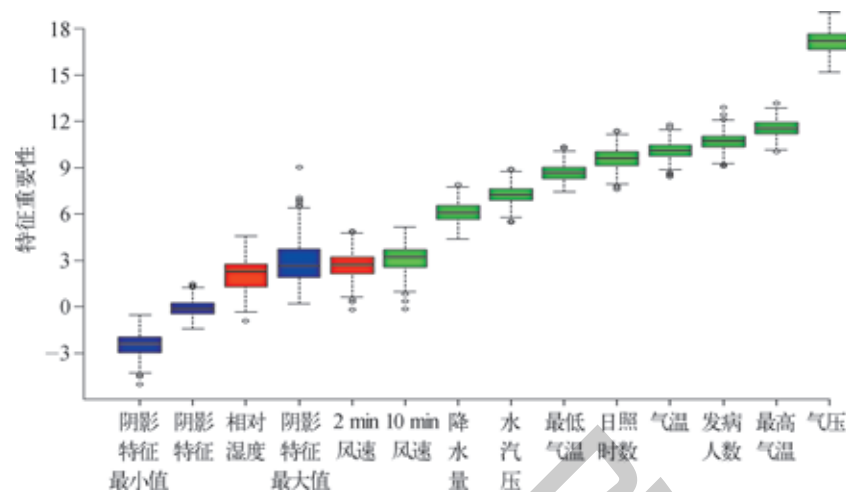
表 2 2009—2016 年重庆市菌痢发病率与气象因子的相关性分析
Tab 2 Correlation analysis of incidence of bacillary dysentery and meteorological factors in Chongqing from 2009 to 2016

气象要素	ρ_{DCCA}			
	均值	中位数	最小值	最大值
月平均水汽压 /hPa	0.857	0.876	0.734	0.886
月平均气温 /℃	0.870	0.903	0.674	0.923
月平均最高气温 /℃	0.875	0.914	0.645	0.938
月平均最低气温 /℃	0.860	0.888	0.693	0.906
月总日照时数 /h	0.741	0.820	0.320	0.893
月总降水量 /mm	0.664	0.729	0.312	0.814
月平均 2 min 风速 /(m/s)	0.551	0.625	0.103	0.760
月平均 10 min 风速 /(m/s)	0.570	0.646	0.120	0.777
月平均湿度 /%	-0.489	-0.549	-0.672	-0.106
月平均气压 /hPa	-0.918	-0.958	-0.974	-0.687

2.4 菌痢发病率预测模型

采用 Boruta 算法进行预测模型特征集合的选择; 结果显示除平均相对湿度和平均 2 min 风速以外, 其余 8 个气象要素被判断为与菌痢发病率具有相关性的重要特征集合 (图 2)。10 个气象要素的 Boruta 算法重要性排序与上述 DCCA 系数法相关性排序几乎相同, 前 3 位分别是平均气压、平均最高气温以及平均气温, 最后 3 位分别是平均

10 min 风速、平均 2 min 风速以及平均相对湿度。Boruta 算法与 DCCA 系数法均可有效筛选出与模型因变量具有相关性的特征变量, 但不同的是 DCCA 系数法量化了相关性的大小及方向。基于以上结果, 本研究将菌痢月发病率作为因变量, 将 8 个重要气象要素特征集合及每月发病率的滞后 1 个月菌痢发病人数作为自变量建立 PSO_SVR 模型。



注：蓝色的箱图代表阴影特征的最小、平均和最大 Z 分数；红色和绿色的箱图分别表示被拒绝和确认为重要特征的 Z 分数。

图 2 Boruta 算法相关性特征集合筛选结果
Fig 2 Boruta algorithm relevance feature set screening results

选择 ε -不敏感函数作为 SVR 损失函数，径向基函数作为 SVR 核函数，采用 PSO 算法寻找最优 SVR 模型惩罚参数 C 和径向基核函数参数 γ ，以此构建 PSO_SVR 模型。在 PSO 算法中将粒子种群最大迭代数、种群数量分别设置为 200、20，加速度因子 c_1 、 c_2 分别设置为 1.5、1.7，以 MSE 作为粒子适应度值。为提高 PSO 算法参数寻优能力，将惯性权重设置为线性递减权重： $w(k)=w_{start}-(w_{start}-w_{end})\times k/T_{max}$ 。式中 w_{start} 为初始惯性权重， w_{end} 为最大迭代次数权重，其值分别设置为 0.9、0.4； k 与 T_{max} 分别为种群当前迭代次数与最大迭代次数；同时，引入变异算子，使得种群粒子在每一次迭代之后，以一定的概率重新初始化粒子，以此改善 PSO 算法容易早熟收敛的缺点^[13]。最终建立的 PSO_SVR 模型的 MSE 、 $MAPE$ 和 R^2 分别为 0.055、0.101、0.909。为了与 PSO_SVR 模型比较，建立了模型自变量只纳入月发病人数的 PSO_SVR1 模型，其 MSE 、 $MAPE$ 和 R^2 分别为 0.129、0.123 与 0.786。进一步建立了没有采用 PSO 优化的

SVR 模型（自变量为菌痢发病人数及气象要素）以及没有采用 PSO 优化的 SVR1 模型（自变量仅为菌痢发病人数），其预测精度详见表 3。结果显示，采用 PSO 算法结合气象要素构建的菌痢发病率预测模型 PSO_SVR 效果最好，其预测性能与 PSO_SVR1、SVR 及 SVR1 模型相比， MSE 分别降低了 57%、52% 和 67%； $MAPE$ 分别降低了 18%、26% 和 45%； R^2 分别提高了 16%、10% 和 28%。最后，绘制 PSO_SVR 在训练集与测试集上的拟合曲线（图 3）。

表 3 菌痢发病率预测模型预测性能比较
Tab 3 Comparison of predictive accuracy of bacterial dysentery incidence prediction models

模型	MSE	$MAPE$	R^2
PSO_SVR	0.055	0.101	0.909
PSO_SVR1	0.129	0.123	0.786
SVR	0.116	0.136	0.827
SVR1	0.166	0.183	0.711

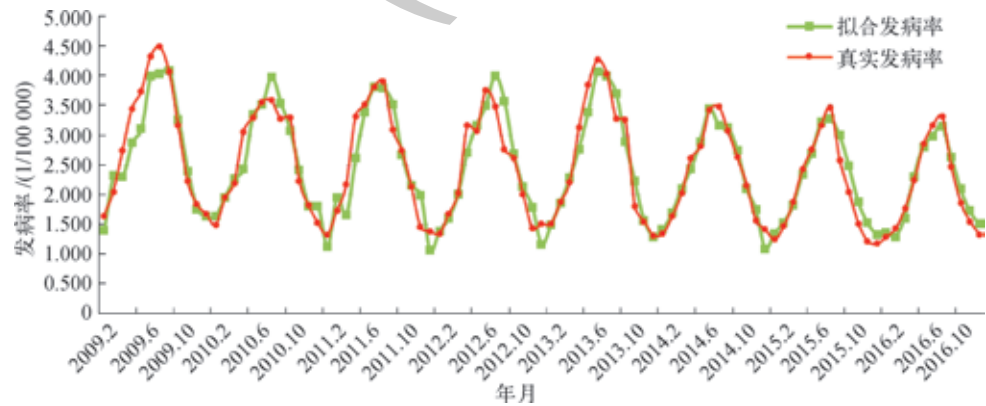


图 3 PSO_SVR1 模型的预测发病率与实际发病率
Fig 3 Predicted incidence and actual incidence of the PSO_SVR1 model

3 讨论

菌痢具有发病率高、传播范围广、季节性强等特点。本研究发现 2009—2016 年重庆市累计报告菌痢 68 855 例, 年均发病率为 29.394/100 000, 发病率虽然总体呈现下降趋势, 但仍远高于全国同期水平 (2009 年 20.28/100 000, 2012 年 15.29/100 000, 2014 年 11.24/100 000)^[2, 14], 这凸显出重庆市菌痢防控的严峻形势。由空间流行病学分析结果可知, 经济较发达、人口流动性与人口密度较大的重庆市主城区及经济相对较落后的渝东北部分地区 (如梁平县、城口县、垫江县等) 应作为防控菌痢流行的重点区域; 0~5 岁人群、5~10 岁人群、散居儿童、农民等应作为重点防控人群, 同时不可忽视的是人群菌痢发病的性别差异性。本研究结果显示气象要素与人群菌痢发病存在强相关性; 其中气压与人群菌痢发病率存在强负相关性, 表明随着气压的降低, 人群菌痢发病率将升高; 气温与人群菌痢发病率存在强正相关性, 表明随着气温的升高, 人群菌痢发病率也将升高; 这符合菌痢夏秋季发病率季节性升高的特点 (夏秋季气温升高, 气压降低), 与流行病学研究结果一致。因此, 相关卫生部门除了关注菌痢流行的重点防控区域、重点防控人群以外, 还应结合气象要素与人群发病率的相关性采取针对性的应对措施, 如可通过大众传播、新媒体等方式进行全方位的健康促进, 普及菌痢防控知识, 以此来控制菌痢的传播与流行。

传染病发病率的预测是其疫情防控的基础, 能为相关卫生部门制定解决方案、降低风险损失提供决策支

持, 在疾病研究领域、流行病防治策略制定上有着重要意义。本研究基于气象要素与人群菌痢发病率的相关性, 运用 Boruta 算法结合 PSO 算法建立了 PSO_SVR 模型。与其他学习方法 (如 ARIMA 模型、BP 神经网络) 相比, SVR 具有更好的泛化能力, 在小样本、非线性问题的模式识别上具有很大的优势^[15-16]。结果显示, 本研究建立的 PSO_SVR 模型 *MSE*、*MAPE*、*R*² 分别达到了 0.055、0.101、0.909, 其预测性能较同类研究结果明显提高。如孙小宇等^[17]分别应用温特斯法模型及多层感知器模型预测细菌性痢疾的月发病率, 其 *MAPE* 分别为 0.266 7 和 0.471 7; 在郑慧敏等^[18]和解合川等^[19]各自的菌痢月发病率预测模型研究中, *MAPE* 分别为 0.296 和 0.132 3。同时, 通过与本研究建立的其他模型对比, 也印证了气象要素作为菌痢发病率模型特征变量的重要性及合理性。

综上, 本研究探讨了 2009—2016 年重庆市菌痢的空间流行特征, 量化了气象要素与人群菌痢发病率的相关性, 并基于此建立了相应的预测模型且取得了良好的预测效果, 这可为重庆市菌痢的防控、卫生资源的合理配置提供理论支撑。此外, 菌痢的发病与流行受到多种因素的影响, 在今后的研究中可探索更多的菌痢影响因素, 并在此基础上对预测方法作进一步研究。同时, 由于本研究侧重于描述重庆市菌痢的空间流行病学特征并建立其高效的发病率预测模型, 暂时没有考虑到各地区气象要素与菌痢发病率之间的相关性, 这需要在在今后的研究中予以完善和补充。

参·考·文·献

- [1] 张楠, 曾德唯, 方勇. 重庆市南岸区 2011 年细菌性痢疾报告及监测结果分析 [J]. 现代预防医学, 2013, 40(19): 3694-3695, 3699.
- [2] 常昭瑞, 孙强正, 裴迎新, 等. 2012 年中国大陆地区细菌性痢疾疫情特点与监测结果分析 [J]. 疾病监测, 2014, 29(7): 528-532.
- [3] 罗玲, 黎明. 湖北省水痘发病情况与气象因素的关系研究 [J]. 中国全科医学, 2018, 21(9): 1114-1117.
- [4] 栾博. 吉林省 2008~2016 年手足口病流行特征分析及时间序列模型的应用 [D]. 长春: 吉林大学, 2017.
- [5] 陈虹汝, 周亮, 李高明, 等. 基于气象因素的手足口病发病风险预测模型 [J]. 第三军医大学学报, 2017, 39(12): 1292-1297.
- [6] Zebende GF. DCCA cross-correlation coefficient: quantifying level of cross-correlation [J]. Phys Stat Mech Appl, 2011, 390(4): 614-618.
- [7] Kulldorff M, Heffernan R, Hartman J, et al. A space-time permutation scan statistic for disease outbreak detection [J]. PLoS Med, 2005, 2(3): e59.
- [8] Vassoler RT, Zebende GF. DCCA cross-correlation coefficient apply in time series of air temperature and air relative humidity [J]. Phys Stat Mech Appl, 2012, 391(7): 2438-2443.
- [9] Kursa MB, Rudnicki WR. Feature selection with the boruta package [J]. J Stat Softw, 2010, 36(11): 1-13.
- [10] Poli R, Kennedy J, Blackwell T. Particle swarm optimization [J]. Swarm Intell, 2007, 1(1): 33-57.
- [11] Cao J, Cui H, Shi H, et al. Big data: a parallel particle swarm optimization-back-propagation neural network algorithm based on mapreduce [J]. PLoS One, 2016, 11(6): e0157551.
- [12] Podobnik B, Grosse I, Horvatić D, et al. Quantifying cross-correlations using local and global detrending approaches [J]. Eur Phys J, 2009, 71(2): 243-250.
- [13] 李钰曼. 改进的 PSO-RBF 神经网络在复杂工业过程中的应用 [D]. 石家庄: 河北科技大学, 2018.
- [14] Chang Z, Zhang J, Lu R, et al. The changing epidemiology of bacillary dysentery and characteristics of antimicrobial resistance of *Shigella* isolated in China from 2004-2014 [J]. BMC Infect Dis, 2016, 16(1): 685.
- [15] Esmaeily H, Tayefi M, Ghayour-Mobarhan M, et al. Comparing three data mining algorithms for identifying the associated risk factors of type 2 diabetes [J]. Iran Biomed J, 2018, 1(4): 133-134.
- [16] Zhang J, Nawata K. A comparative study on predicting influenza outbreaks [J]. Biosci Trends, 2017, 11(5): 533-541.
- [17] 孙小宇, 初艳慧, 张震, 等. 3 种预测模型在北京市西城区痢疾发病数预测中的应用与比较 [J]. 现代预防医学, 2014, 41(19): 3470-3474.
- [18] 郑慧敏, 薛允莲, 黄燕飞, 等. ARIMA 模型在深圳市法定传染病发病趋势预测的应用 [J]. 实用预防医学, 2016, 23(02): 240-243.
- [19] 解合川, 任钦, 曾海燕, 等. 支持向量机在传染病发病率预测中的应用 [J]. 现代预防医学, 2013, 40(22): 4105-4108, 4112.

[收稿日期] 2018-10-10

[本文编辑] 吴 洋

