

组合预测模型在女性呼气高峰流量参考值 地理分布研究中的应用

薛然尹,葛 森,何进伟,胡燕宇,谷琳琳,杨绍芳
(陕西师范大学旅游与环境学院健康地理研究所,西安 710119)

摘 要:随着地理学的发展以及人们对健康问题的日益关注,医学地理学得到迅速发展。鉴于目前医学参考值制定时仍存在忽略地理因素的影响,本文收集中国各地3809例健康成年女性呼气高峰流量参考值,分析地理因素对其的影响,计算不同地区参考值的数值差异,探究地理因素对医学参考值产生影响的机理,其中,纬度、海拔高度、年平均气温、年平均相对湿度、年降水量、表土砂砾百分率、表土参考容量共7项地理因素存在显著的相关性。利用ArcGIS中的Moran's I 指数对数据进行分析,确定数据与空间及地理因素存在关系。并通过岭回归分析,建立回归方程,并进行插值。研究结果表明,中国健康成年女性肺部呼气高峰流量与纬度与海拔、气候、土壤等地理因素之间存在着显著的关系,同时证明,岭回归与支持向量机组合模型的地理分布差异预测结果优于单独预测方法。

关 键 词:呼气高峰流量;岭回归;支持向量机;组合预测模型;地理分布;中国

doi: 10.11820/dlkxjz.2014.05.005

中图分类号: R188

文献标识码: A

1 引言

呼气高峰流量(Peak Expiratory Flow Rate, PEFR),又称最大呼气流量(Peak Expiratory Flow, PEF)或呼气流速峰值,指在测定用力肺活量(FVC)过程中的最大呼气流速。国家哮喘教育计划(National Asthma Education Program, NAEP)认为,PEFR测定的目的是:①用以评价哮喘的严重程度,辅助确定治疗方案;②监测急性恶化时肺部对于治疗的反应;③监测进行维持治疗时,慢性哮喘的反应,提供改变治疗方法的客观指标;④监测无症状哮喘者的肺功能变化,以便能在病症急性恶化前进行早期治疗;⑤诊断运动性哮喘。研究发现,呼气肌的收缩力、肺、胸廓的弹性回缩力及大气道阻力是影响最大呼气峰值流速的主要因素,并且还与气管、支气管的横截面密切相关。因此,测定PEFR值能很好的反映出呼气肌、胸廓、气道阻力及气管、支气管是否发生病变。

中国所用的PEFR正常参考值为广州呼吸病研究所20世纪90年代初的人群调查所得。目前对该指标的科学研究以及参考值制定过程中,又多以性别、年龄、身高作为主要影响因素(穆魁津, 1990),倾向于生理因素差异。由于未考虑地理因素对呼气高峰流量的影响,导致制定呼气高峰流量值缺乏统一的标准。有关呼气高峰流量与地理因素之间关系的研究则多见于与海拔之间的关系(王涛等, 2000),以及与气候和空气质量之间的关系(刘英等, 2008; 张晓明等, 1995)。经过数据调查,在排除性别、年龄、身高的影响后,不同地区的参考值仍会具有统计学上的显著差异。

事实上,呼气高峰流量值与人类生活的地理环境关系非常密切,地理环境对呼气高峰流量的影响可通过人们所处的地形地貌、大气条件、土壤条件等自然环境要素,以及饮食结构、居住条件、经济状况等人文环境要素,综合作用于人体呼吸器官的反映。与此同时,对于传统医学研究常用的回归分析方法,要

收稿日期:2013-10;修订日期:2013-12。

基金项目:国家自然科学基金项目(40971060)。

作者简介:薛然尹(1988-),男,河北承德人,硕士生,主要研究方向为健康地理学及GIS分析方法,E-mail: lyws2007@qq.com。

通讯作者:葛森(1960-),男,研究员,博士生导师,主要从事健康地理研究,E-mail: gemiao@snnu.edu.cn

求各分析自变量彼此间无线性相关,但由于各地理要素之间具有较大的相关性,会对模型的模拟预测精度造成影响(丁圣彦等, 2007; 杨青生等, 2006)。因此有必要进一步研究两者之间的复杂关系。

2 数据与方法

2.1 数据来源

2.1.1 呼气高峰流量值数据来源

通过查询中国期刊网数据库、万方数据资源系统、中国生物医学文献数据库、超星数字图书馆、维普中文科技期刊数据库等数据库,并实际测定部分数据,共收集了中国54个市(县)级医院、有关单位及高等院校测定的3809例健康成年女性呼气高峰流量值,数据来源见参考文献(仅列举部分重要数据)(陈莉等, 2004; 邓杰, 1993; 李仲铭等, 2009, 2012; 林江涛等, 1989; 田广智等, 1981; 王兆维等, 1984; 王瑶等, 1988; 徐风珍等, 1985; 谢美云等, 2012; 赵立等, 2009; 赵蓉雅, 2010)。选取其中44组作为数据组,其余10组作为检验组。

由于对数据质量的要求较高,需要尽量排除年龄、仪器等因素对于指标的影响,数据获取较为困难。因此,选取年龄介于23~50岁之间的健康女性来排除年龄影响;测量呼气高峰流量的仪器选用日产Chest STM-81肺功能检查仪以及相近功能的肺功能检查仪。

2.1.2 地理指标选取及数据来源

人体健康与地理因素之间存在着复杂的关系。早在公元前,希腊古典时期著名医生希波克拉底(Hippocrates of Kos)就有关于健康与居住生活的“地方”有关的论述。中国古代医学巨著《黄帝内经·素问》中的至真要大论篇中更有“夫百病之生也,皆生于风寒暑湿燥火,以之化之变也”的论述。在当代,人体健康与地理因素之间存在关系已成为公认的事实。中国医学地理学家谭见安等在《地球环境与健康》一书中明确提出“人类疾病和健康状况与地球环境的性质和变化密切相关……”。单就肺组织来说,肺位于胸腔,左右各一,由于处于五脏最高位置,故而古人称其为“华盖”,有“外输精于皮毛,内通调于四渎,沛然施于周身”的作用;又因肺叶娇嫩,不耐寒热,易被邪侵,又有“娇藏”之称,是人体的“魄之处”、“气之本”(江涛, 2006)。从中可

知,肺组织十分容易受到周围环境的影响。

本文主要选取三类地理因素进行分析:地势指标、气象指标、土壤指标。地势资料摘自有关地理著作和辞典(阎崇年等, 1991; 颜宏等, 2002; 赵济等, 1999);海拔高度与气象指标来自中国气象科学数据共享服务网(<http://cdc.cma.gov.cn/>);土壤指标来自数据来源于联合国粮农组织(FAO)和维也纳国际应用系统研究所(IIASA)所构建的世界和谐土壤数据库(Harmonized World Soil Database)(HWSD)(<http://www.fao.org/nr/land/soils/harmonized-world-soil-database/zh/>)。

选取的地理指标有:经度(X_1),纬度(X_2),海拔高度(X_3 , m),年日照时数(X_4 , h),年平均气温(X_5 , $^{\circ}\text{C}$),年平均相对湿度(X_6 , %),年降水量(X_7 , mm),气温年较差(X_8 , $^{\circ}\text{C}$),年平均风速(X_9 , m/s),表土砂粒含量(X_{10} , % wt.),表土粉粒百分率(X_{11} , % wt.),表土黏粒百分率(X_{12} , % wt.),表土参考容重(X_{13} , kg/dm^3),表土容重(X_{14} , kg/dm^3),表土石砾含量(X_{15} , % vol.),表土有机质含量(X_{16} , % wt.),表土pH值(X_{17}),表土阳离子交换量(粘土) (X_{18} , cmol/kg),表土阳离子交换量(潜育土) (X_{19} , cmol/kg),表土盐基饱和度(X_{20} , % wt.),表土总可交换量(X_{21} , $\text{cmol}/\text{kg wt.}$),表土碳酸钙含量(X_{22} , % wt.),表土硫酸钙含量(X_{23} , % wt.),表土碱度(X_{24} , % wt.),表土盐分(X_{25} , dS/m)共25项。

2.2 分析及预测方法

2.2.1 相关分析

用SPSS软件构建中国健康成年女性呼气流速峰值参考值与地理因素数据库,并进行相关分析,提取出具有显著相关性的地理因素。

2.2.2 空间自相关

空间自相关是指同一个变量在不同空间位置上的相关性,是空间单元属性值聚集程度的一种度量(Getis et al, 1992)。通过对采样数据进行全局空间自相关检验,如能证明所测指标数据具有空间相关性,则反映地理区域的差异会对医学指标产生不可忽略的影响。

2.2.3 岭回归分析

岭回归分析(Ridge Regression)是一种用来解决多重共线性的有偏估计回归方法,实际上是一种放弃无偏性的改良最小二乘法,以损失部分信息和放弃部分精确度为代价,求得效果稍差但是更符合实际的回归方程,它对病态数据的耐受性要远远强

于最小二乘法。岭回归分析非常灵活,它的使用过程中会存在一定的主观人为性,但这正是发挥定性分析与定量分析有机结合之处,在解决多重共线性问题中可发挥独特作用(杨楠, 2004)。

2.2.4 支持向量回归机(SVR)

应用在解决小样本、非线性及高维模式识别中表现出特有优势的支持向量回归机(Support Vector Regression, SVR)进行回归分析建模。函数回归问题可以描述为如何寻找一个函数,使得期望风险最小,实际情况中多倾向于非线性关系。运用SVR将每一个样本点用一个非线性函数 Φ 映射到高维特征空间,再在高维特征空间进行线性回归,从而取得在原空间非线性回归的效果(夏国恩等, 2006)。函数 f 为:

$$f(x)=w \cdot \Phi(x)+b(x_i, y_i), (i=1,2,\dots,n) \tag{1}$$

式中: $x_i \in R^n$ 为输入量; $y_i \in R$ 为输出量; n 为数据数量; Φ 为非线性函数; b 为常数,需要确定 Φ 和 b 。

这种算法使得寻优函数在高维空间时间上只需进行内积运算,而SVR的内积运算是可以用原来空间中的函数实现的,甚至没有必要知道变换的形式,因此可以直接输出拟合值。

2.2.5 组合预测模型

在1969年Bates和Granger的论文“组合预测”中,首次提出模型综合研究方法,并有效地应用于模拟预测。在经历一系列理论研究和实际应用后,组合预测模型表现出比单个预测模型具有更高的预测精度,能增强预测的稳定性,与单个模型的长处相结合,具有较高的适应未来预测环境变化的能力。所谓组合预测方法就是通过一定的方法对多个预测结果进行综合,以得到一个较窄的预测值取值范围供系统分析或决策用。常用方法大致可以分为权重合成法和区域合成法。

区域合成法要求至少有3个模型,因此选取权重合成法。权重合成法又可分为平均值法、标准差法、二项式系数法、离异系数法、AHP法等,基于所选预测模型最终确定标准差法最为合适。其权重系数 w_i 为:

$$w_i = \frac{s - s_i}{s} \cdot \frac{1}{n - 1}; s = \sum_{i=1}^r s_i (i=1, \dots, r) \tag{2}$$

式中: s_i 为第 i 个模型的标准差(孙林岩, 1990), n 为模型个数。

2.2.6 空间差异预测图绘制

为了准确揭示出中国成年女性呼气高峰流量

参考值的地理分布规律,选取全国2322个市县作为观测点,通过模型计算出2322个市县成年女性呼气高峰流量参考值的预测值。应用ArcGIS软件中的空间分析模块进行克里金(Kriging)插值(汤国安等, 2006),并输出空间预测图。

3 结果与分析

3.1 原始数据统计及相关分析

经整理和筛选后的数据有54组,共3809例,平均年龄35.08岁,呼气高峰流量平均值为5.43 L/s。这其中,选取44组作为数据组,其余10组作为对照组。数据空间分布尽量选取地理因素典型及具有代表性的市(县)。

运用相关分析,计算出中国成年女性呼气高峰流量参考值与上述25项地理指标的单相关系数(r)和 P 值(表1)。

由表1中的相关分析可以看出,选取的25个地理因素指标中,有7个地理因素与呼气高峰流量值具有相关性,分别是:纬度、海拔高度、年平均气温、年平均相对湿度、年降水量、表土砂砾含量、表土参考容重。其中,海拔高度、年平均气温、年降水量与呼气高峰流量值具有极其显著的相关性。

3.2 空间自相关分析

利用Arcgis中的Spatial Statistics Tools中的空间自相关模块,对实测数据点的呼气高峰流量值进行全局空间自相关分析(徐建华, 2006),结果如表2所示。

通过计算,得出全局自相关系数, Z 值为2.076476,大于0.05的置信水平的1.96,有很强的空间自相关性。因此,中国成年女性呼气高峰流量参考值与地理环境并不是相互独立的,存在着相关性,空间属性的差异会造成参考值的变化。

3.3 岭回归分析

将25项地理因素中提取出与呼气高峰流量相关的纬度、海拔高度、年平均气温、年平均相对湿度、年降水量、表土砂砾含量、表土参考容重共7项地理因素进行相关分析,分析结果见表3。

由表3可知,各地理因素间存在多重共线性,如 X_2 与 X_5 之间相关系数为-0.912, X_2 与 X_7 之间系数为-0.879, X_{10} 与 X_{13} 之间系数为-0.861,等等。故消除共线性是十分必要的,适合于岭回归分析。利用SAS软件输出岭迹图(图1)。

观察岭迹图可知,各回归系数在岭参数 $k=0.2$ 时趋于稳定,且方差膨胀因子均小于1,此时岭回归估计的残差标准差 $RMSE=0.84628$,虽比最小二乘回归方程的残差标准差(0.82071)有所增大,但基本可以忽略。因此取 $k=0.2$,计算出中国成年女性呼气高峰流量参考值与地理因素之间的回归方程:

表1 呼气高峰流量实测值与地理因素相关系数
Tab.1 Correlation coefficients between $PEFR$ measured value and geographical factors

地理因素	相关系数 r	P 值	地理因素	相关系数 r	P 值
X_1	-0.166	0.229	X_{14}	-0.221	0.108
X_2	0.288*	0.035	X_{15}	0.169	0.221
X_3	0.423**	0.001	X_{16}	0.115	0.408
X_4	0.118	0.396	X_{17}	0.005	0.973
X_5	-0.405**	0.002	X_{18}	-0.191	0.167
X_6	-0.295*	0.031	X_{19}	0.130	0.350
X_7	-0.388**	0.004	X_{20}	-0.055	0.690
X_8	0.120	0.388	X_{21}	0.178	0.198
X_9	-0.054	0.699	X_{22}	0.146	0.291
X_{10}	-0.320*	0.018	X_{23}	0.045	0.745
X_{11}	0.201	0.144	X_{24}	0.000	0.998
X_{12}	0.210	0.128	X_{25}	0.037	0.791
X_{13}	-0.305*	0.025			

注:**表示在0.01水平(双侧)上显著相关;*表示在0.05水平(双侧)上显著相关。

表2 Moran 报表
Tab.2 Moran's report

全局 Moran's I 汇总				
Moran's I 指数	预期指数	方差	z 得分	P 值
0.265868	-0.037037	0.021279	2.076476	0.037850

$$\hat{Y}_1=10.1007-0.0061X_2+0.0002X_3-0.0075X_5-0.00832X_6-0.0005X_7-0.01789X_{10}-1.9516X_{13}\pm0.84628$$

(3)

对比回归方程(3)与表1发现,岭回归方程表现出了各地理因素与 $PEFR$ 之间的关系。但需要注意的是, X_2 的相关系数为正,而在方程中的系数为负,这是由于分析因子的多重共线性所致,若取岭参数 k 大于0.4,则 X_2 在方程中的系数可变为正,但由于残差标准差增大过多,会导致模型拟合效果很差,因此仍选用 $k=0.2$ 时拟合出的方程。

3.4 支持向量回归机预测

目前关于SVM的算法以及应用已经比较成熟,以LIBSVM 3.1 软件包(Faruto, 2010)为基础进行Matlab 编程设计。在预测型学习任务中,模型及其参数选择是有效预测的前提(刘倩等, 2011)。采用 ϵ -SVR(ϵ : epsilon, 不敏感损失函数中的偏差)作为预测模型,核函数为RBF(径向基)核函数。

首先对数据进行归一化预处理,为防止数据泛化,需要进行降维预处理(图2),可以观察到经过降

表3 地理因素相关系数表
Tab.3 Correlation coefficients of geographical factors

	X_2	X_3	X_5	X_6	X_7	X_{10}	X_{13}
X_2	1.000	0.408	-0.912	-0.806	-0.879	0.226	0.190
X_3	0.408	1.000	-0.680	-0.618	-0.563	-0.042	-0.108
X_5	-0.912	-0.680	1.000	0.814	0.823	-0.039	0.017
X_6	-0.806	-0.618	0.814	1.000	0.845	-0.298	-0.320
X_7	-0.879	-0.563	0.823	0.845	1.000	-0.235	-0.297
X_{10}	0.226	-0.042	-0.039	-0.298	-0.235	1.000	0.861
X_{13}	0.190	-0.108	0.017	-0.320	-0.297	0.861	1.000

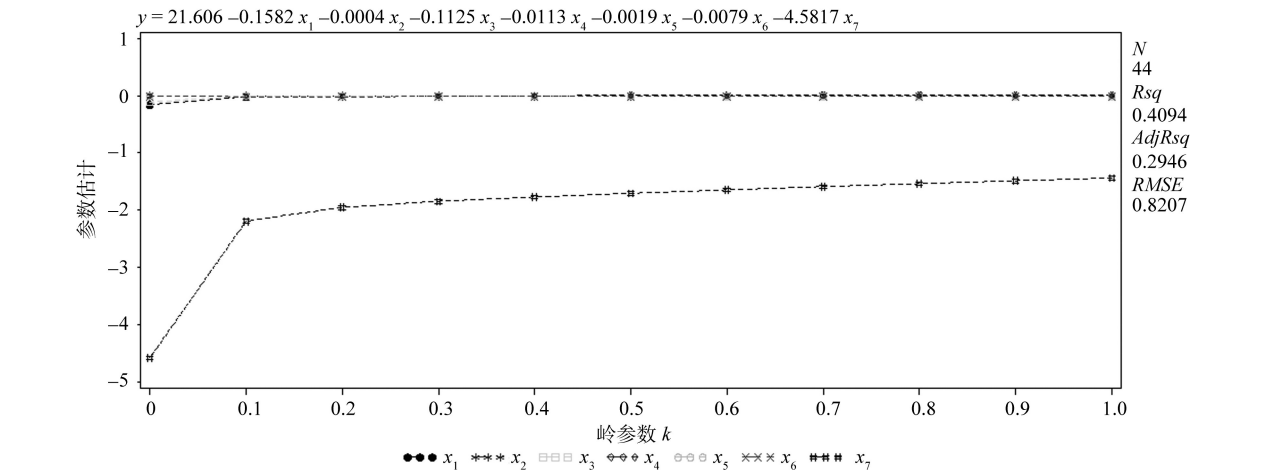


图1 地理因素岭迹图
Fig.1 Ridge trace plot of geographical factors

维预处理的因子已由原有的7个因子减少到4个因子,此时的方差贡献率已达到95%以上。

降维之后,将44个实测数据组作为训练数据,10个检验数据组作为检验数据寻找最优参数,命令如下: [bestmse, bestc, bestg] =SVMcgForRegress (train-y-scale, train-x, -3, 3, -3, 3, 10, 0.2, 0.2, 0.2)。惩罚参数 c 的范围是 $[2^{-3}, 2^3]$,RBF核参数 g 的范围是 $[2^{-3}, 2^3]$,Cross Validation过程中的参数为10折,参数寻优 c 和 g 的步进大小为0.2,准确率离散化显示的步进间隔为0.2。最终寻得最优参数 $c=1.5157, g=2.639$,训练误差 $mse=0.04107$ 。具体结果见图3,拟合程度以及拟合值与实际值的比较见图4。

由图4可以观察出拟合结果比较理想,可进一步的预测。将所有预测数据导入程序,得到最终预测数据,预测值分布见图5。

3.5 组合模型预测

采用标准差法体现了按模型拟合精度加权综合的思想,能够较好的将模型预测结果拟合度加强。为计算权重,首先计算标准差,利用SPSS进行双侧 T 检验,得出岭回归分析的预测结果与真实值的标准差为0.7546,支持向量回归机预测结果与真实值的标准差为0.7285。利用公式(2)对岭回归分析和支持向量回归机的预测结果进行权重计算,计算结果 $\omega_1=0.4912, \omega_2=0.5088$ 。构建组合预测模型,模型预测效果见表4。可以看到,岭回归分析标准差值大于支持向量机标准差值。仅从这10组检验数据的标准差和 P 值来看,岭回归模型差于支持向量机模型。但10组检验数据仅能反映预测的大致趋势,经过组合预测后,模型标准差值基本与支持

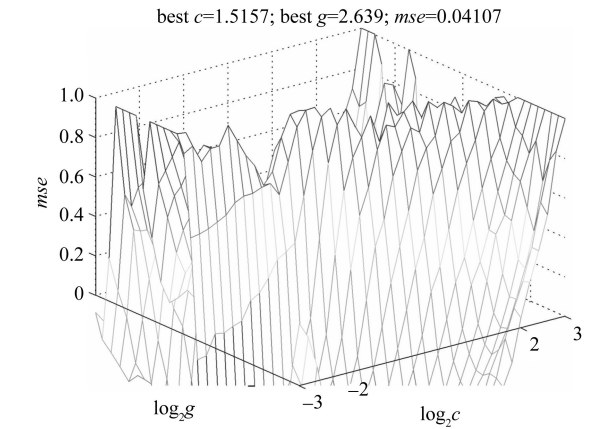


图3 SVR参数寻优
Fig.3 SVR parameter optimization

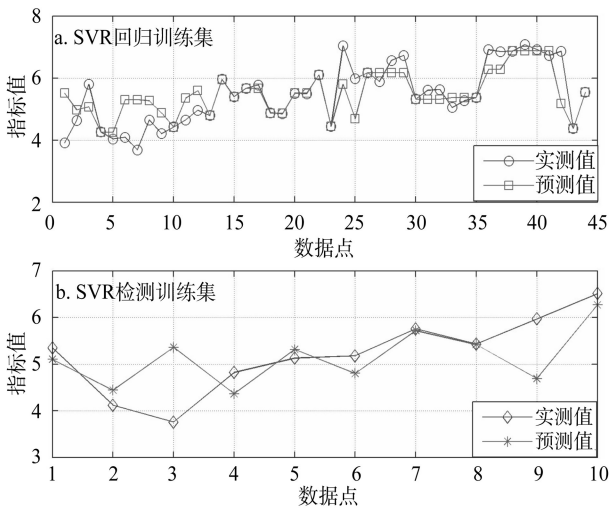


图4 模型预测拟合结果
Fig.4 Fitting results of model prediction

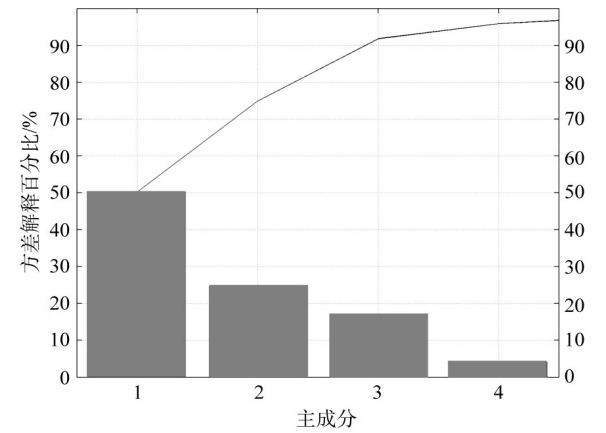


图2 降维图
Fig.2 Dimensionality reduction plot

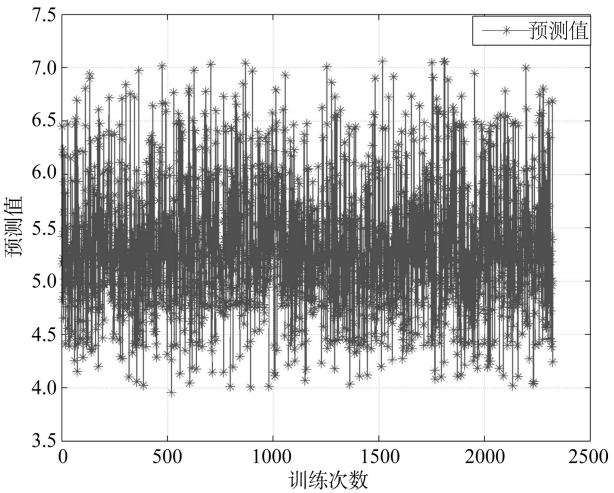


图5 预测值分布图
Fig.5 Distribution of predicted values

向量机模型持平,而且能够突出两个模型的预测特点。因此,组合预测的结果是成功的。

3.6 呼气高峰流量参考值空间差异预测

为直观揭示中国成年女性呼气高峰流量参考值的地理分布规律,需要用ArcGIS软件精确拟合中国成年女性呼气高峰流量参考值的空间预测图。首先对组合预测模型的预测数据以SPSS软件进行K-S检验(验证数据是否为正态分布),检验结果Z值为1.671,渐近显著性为0.07,属于正态分布。因此插值方法宜采用普通克里金法(徐芝英等, 2012)。

利用ArcGIS地统计模块中的析取克里金法,采用块金值(Nugget, 0.72737)和偏基台(Partial Sill, 0.27059)都比较合适的协方差函数进行建模(陈春阳等, 2012),建模效果如图6所示。最终输出插值图(图7),在图中标注了部分城市和地区的中国成年女性呼气高峰流量参考值,位于同一颜色的区域呼气高峰流量参考值相等,颜色相近的区域参考值相近,颜色变化多的区域表示参考值变化较快,反之变化缓慢。

由图7可知,中国成年女性呼气高峰流量参考值整体呈现西北高、东南低的趋势;同时表现出一定的纬度正相关性,即随纬度增加正常参考值随之

增加;与海拔高度也体现出了极其显著的相关性;沿海地区相较其临近的陆地地区正常参考值偏小。

PEFR值的地区差异方面,位于西部的西藏自治区以及青海地区的PEFR值较高;高值区域集中在内蒙古,陕西、青海、甘肃、吉林等地;沿海地区的广西、广东、江苏等省区沿海地区的PEFR值较低,内陆地区该值偏小。PEFR值最小地区出现在浙江附近;最大值则出现在内蒙古、甘肃地区。

4 讨论与结论

4.1 分析与讨论

已有研究表明,成年女性呼气高峰流量值的大小直接取决于以下因素:大、小气道的通畅程度,肺组织的弹性,胸廓的弹性,呼吸肌的力量,以及受测者的配合程度。因而,会明显受到年龄、身高、性别的影响,另外也受测试季节和时段的影响(赵瑛, 1999; 赵海燕等, 2003)。而本文研究结果显示,中国成年女性呼气高峰流量参考值与7项地理因素具有相关性,分别是:纬度、海拔高度、年平均气温、年平均相对湿度、年降水量、表土砂砾含量、表土参考容重。其中,有3项指标显著性较强,为影响PE-

表4 模型效果比较
Tab.4 Model comparison

城市	上海市	广州市	南京市	北京市	杭州市	诸暨市	开封市	舟山市	重庆市	西安市	标准差	P值
岭回归	5.21	4.86	5.41	4.43	5.16	4.91	5.54	5.52	4.95	5.79	0.7546	0.923
支持向量机	5.10	4.44	5.36	4.36	5.31	4.80	5.72	5.42	4.69	6.27	0.7285	0.815
组合模型	5.16	4.65	5.38	4.39	5.24	4.85	5.63	5.47	4.82	6.03	0.7309	0.867
真实值	5.35	4.12	3.76	4.82	5.13	5.18	5.76	5.43	5.97	6.52		

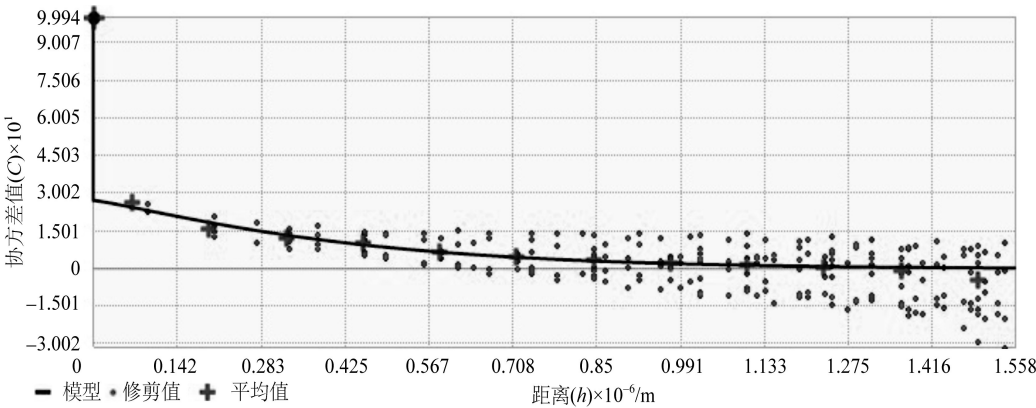


图6 协方差模型效果
Fig.6 Results of covariance model

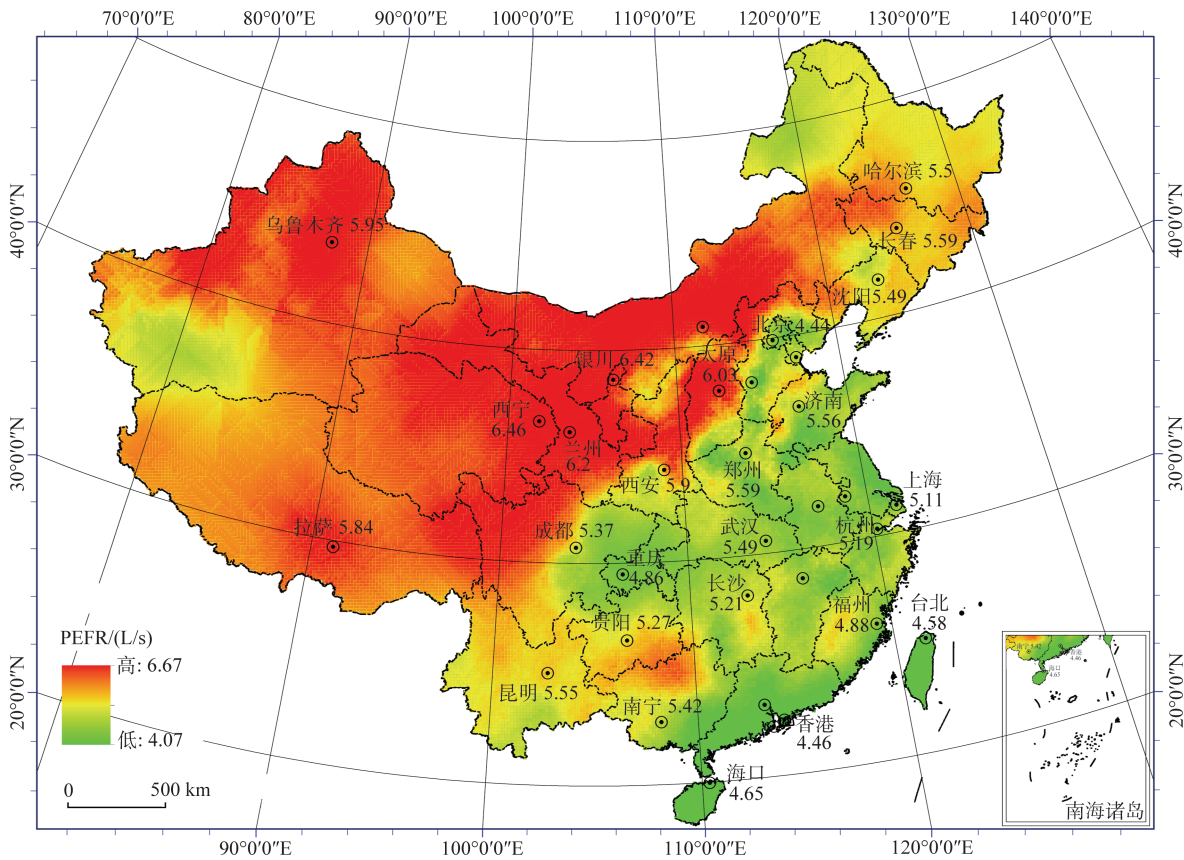


图7 中国成年女性呼气高峰流量参考值空间差异预测图

Fig.7 Spatial difference prediction of Chinese adult women's PEFR

FR 参考值的主要影响因素,分别是海拔(显著性水平0.001)、年平均气温(0.002)和年降水量(0.004)。

4.1.1 纬度与海拔影响因素

纬度因素与PEFR值呈显著正相关,这是因为随着纬度的增加,人群平均身高呈增高趋势。高纬度地区终年寒冷,高大的身体对应的比表面积(表面积/体重)小,散热少,利于抵御风寒。而医学研究已经证明,PEFR值与身高的关系是极其显著的正相关,因此PEFR正常参考值会随着纬度的增加而增加。同时,也从侧面证明了温度与PEFR值有一定的关系。

海拔高度与PEFR值的关系是PEFR研究的重点。随着海拔的升高,大气压强逐步降低(宋辞等, 2012),随着空气密度下降,从而使气道、尤其是支气管发生扩张,加快气流速度。同时,空气密度下降造成氧含量降低,故需要更大的吸气,这也造成弹性回缩力的增大,进而导致呼气流速的增加长期生活于高海拔环境的居民,呼吸肌的力量势必强于低海拔环境居民,造成其呼气高峰流量参考值较低

海拔地区高(杨生岳等, 1991)。

考虑到气压与呼气高峰流量值的季节变化具有规律性,我们需要进一步考虑到这两者时间尺度或者空间尺度的关系。空间尺度上,中国南部处于副热带高压带而北部处于盛行西风带,气压呈现南高北低的趋势,而由图7可知,呼气高峰流量值呈现的是北高南低的趋势。时间尺度上,中国大陆整体气压趋势为秋冬高,春夏低。人体肺部的各种病症,都普遍呈现出春末秋初加重的现象,即中医所谓“肺应秋”现象。因此,基于上述论述,可以认为呼气高峰流量值与海拔和气压有着密切的关系。

4.1.2 气候影响因素

由上文相关分析可知,选取的年平均气温、年平均相对湿度、年降水量均与呼气高峰流量呈负相关性。当温度偏低时,空气密度增大,胸内压增大;同时为保证体内温度恒定,新陈代谢速率会增加以维持足够的热量,故而导致参考值增大。需要指出的是,在对数据进行处理时,发现年平均气温的指标对儿童与成人PEFR指标的相关性是相反的。我

们推断这是由于气候环境对人体影响的持续性以及儿童与成人体内新陈代谢差异所共同导致的。

大气湿度是水汽的表征量,人体处于相对湿度40%~60%的环境中最为舒适,气温处于适中情况下,湿度的变化对人体作用较小(毛飞等, 2011),但高温高湿或者短期内潮湿与干燥气候频繁交替对身体机能极其不利,致病风险也会上升。

4.1.3 土壤影响因素

不同的土壤质地,会对其所属区域的小气候、空气质量、生态环境等方面施加影响,间接作用于人体(陈晓琳等, 2011);同时,土壤中的基本元素和有毒物质可通过食物和水,以及呼吸作用摄入人体,直接对人体造成影响(Hallberg等, 2007)。

相关分析结果显示,呼气高峰流量与表土砂砾含量、表土参考容重显著负相关。土壤砂砾含量体现的是砂砾在土壤成分中所占的比例。砂砾含量越高,土壤越倾向于砂质。土壤容重又称土壤密度,是干的土壤基质物质的量与总容积之比,土壤越疏松多孔,容重越小;土壤越紧实,容重越大。总体来说,这两个指标数值越大,土壤质地越差(汤青等, 2013)。当土壤砂砾含量和表土参考容重偏大时,分析区域空气中的颗粒物浓度偏大,可吸入颗粒物含量整体就会增大。一般而言,这两个指标偏高地区多为干旱、半干旱地区,居民长期处于沙尘影响之下。大气中不同粒径颗粒物,尤其是可吸入颗粒物浓度偏高时,会影响人体皮肤、眼、呼吸系统及循环系统健康(Leuenberger et al, 1998; 王金玉等, 2013)。这种长期影响势必会对人体健康指标常量造成影响。

4.2 总结与展望

(1) 利用Moran's I 指数对医学指标与空间关系进行分析,在当前医学领域相关研究中是较少见的,其原因在于地理学与医学工作者专业知识间存在结构差异性,导致彼此方法难以融合。传统的仅考虑性别、身高、体重等指标的医学统计方法存在一定的不完整性,综合考虑各种因素所带来的影响才是未来研究的发展方向。用Moran's I 指数对医学指标进行空间自相关分析,可以直接确定医学指标是否会受到空间差异的影响,可为各种医学指标的生理、病理学研究提供新思路,以及简单快捷的验证方法。

(2) 组合预测模型在其他领域诸如经济、人文地理等领域早有较成熟的应用,且多应用于时间预测建模,在医学地理方面至今应用仍较少。本文力图

验证组合预测模型在医学地理学及空间预测中的可行性,但由于数据质量不十分理想,组合预测模型的结果未达到预期的准确度。这是因为组合预测模型一般应用于3种及3种以上模型的综合和预测,仅有2种预测模型无法利用诸如区域合成法、二项式系数法、离异系数法等预测模型进行对比分析。

(3) 未来对于各影响因子的作用机理探究仍有待加强。尽管地理因素与人体健康之间的复杂关系制约了机理研究的进一步深入,但随着近年来医学地质学、医学地理学的兴起,未来这一方向可望涌现出更多的代表性成果。

参考文献(References)

- 陈春阳, 戴君虎, 王焕炯, 等. 2012. 基于土地利用数据集的三江源地区生态系统服务价值变化. 地理科学进展, 31(7): 970-977. [Chen C Y, Dai J H, Wang H J, et al. 2012. Changes of the value of ecosystem services in the Sanjiangyuan Region based on land use data. Progress in Geography, 31(7): 970-977.]
- 陈莉, 赵明, 韩少梅, 等. 2004. 河北省正常人群肺功能检测分析. 中国医学科学院学报, 26(4): 463-466. [Chen L, Zhao M, Han S M, et al. 2004. Testing and analyzing the lung functions in the normal population in Hebei Province. Acta Academiae Medicinae Sinicae, 26(4): 463-466.]
- 陈晓琳, 李忠武, 王晓燕, 等. 2011. 中亚热带红壤丘陵区松林生态系统表层土壤活性有机碳空间分异规律. 地理研究, 30(10): 1825-1834. [Chen X L, Li Z W, Wang X Y, et al. 2011. Spatial distribution of topsoil labile organic carbon of pinus elliottii ecosystem in the central subtropical region. Geographical Research, 30(10): 1825-1834.]
- 邓杰. 1993. 呼气的最大峰流速对游泳运动员呼吸肌功能的评价意义初探. 四川体育科学, (增刊): 10-12. [Deng J. 1993. The evaluative meaning of peak expiratory flow rate on swimmer respiratory muscle function. Sichuan Sports Science, (Suppl): 10-12.]
- 丁圣彦, 梁国付. 2007. 地理环境因素对伊洛河流域森林景观的影响. 地理研究, 26(5): 906-914, 1071. [Ding S Y, Liang G F. 2007. Analysis of geographic environmental factors on forest landscape dynamics of Yiluo River Basin. Geographical Research, 26(5): 906-914, 1071.]
- Hallberg R, 王胜. 2007. 医学地质学论文前言. AMBIO-人类环境杂志, 36(1): 62, 114. [Hallberg R, Wang S. 2007. Preface to medical geology papers. AMBIO- Journal of Human Environment, 36(1): 62, 114.]
- 江涛. 2006. "肺主气"的文献研究及肺亦为后天之本论[D]. 济南: 山东中医药大学. [Jiang T. 2006. Literature study

- on the theory of "the lung dominating qi" and discussion of "the lung being the coordinative acquired foundation" [D]. Jinan, China: Shandong University of Traditional Chinese Medicine.]
- 李仲铭, 李静华, 白永, 等. 2012. 内陆城市与海岛居民健康成人肺功能比较研究. 昆明医科大学学报, (6): 56-61. [Li Z M, Li J H, Bai Y, et al. 2012. Comparison study on the lung function of people in inland city and island. Journal of Kunming Medical University, (6): 56-61.]
- 李仲铭, 李静华, 徐成丽, 等. 2009. 浙江省4地区正常成年人肺功能检测分析. 现代预防医学, 36(17): 3310-3312, 3315. [Li Z M, Li J H, Xu C L, et al. 2009. Testing and analyzing the lung function of normal adults at four areas in Zhejiang Province. Modern Preventive Medicine, 36(17): 3310-3312, 3315.]
- 林江涛, 林友华, 薛福林. 1989. 272例健康人最大呼气流量—容积曲线测定及年龄对通气功能的影响. 中日友好医院学报, 3(3): 164-168. [Lin J T, Lin Y H, Xue F L. 1989. Measurement of maximum expiratory flow-volume curve in 272 healthy subjects and influence of aging on ventilatory function. Journal of China-Japan Friendship Hospital, 3(3): 164-168.]
- 刘倩, 刘三阳, 葛森. 2011. 基于支持向量机的青年女性红细胞压积参考值与地理环境关系研究. 工程数学学报, 28(1): 1-6. [Liu Q, Liu S Y, Ge M. 2011. Research on relationship between reference value of young women's hematocrit and Chinese geographical factors based on support vector machines. Chinese Journal of Engineering Mathematics, 28(1): 1-6.]
- 刘英, Kim Dae-seon, 何长杰, 等. 2008. 包头市小学生最大呼气流速与扬沙天气关系的分析. 环境与健康杂志, 25(1): 7-10, 95. [Liu Y, Dae-seon K, He C J, et al. 2008. Correlation between sand dust weather and peak expiratory flow rate of children in Inner Mongolia of China. Journal of Environment and Health, 25(1): 7-10, 95.]
- 毛飞, 孙涵, 杨红龙. 2011. 干湿气候区划研究进展. 地理科学进展, 30(1): 17-26. [Mao F, Sun H, Yang H L. 2011. Research progress in dry/wet climate zoning. Progress in Geography, 30(1): 17-26.]
- 穆魁津. 1990. 全国肺功能正常值汇编. 北京: 北京医科大学出版社. [Mu K J. 1990. The normal value of pulmonary function in China. Beijing, China: Beijing Medical University Press.]
- 宋辞, 裴韬, 周成虎. 2012. 1960年以来青藏高原气温变化研究进展. 地理科学进展, 31(11): 1503-1509. [Song C, Pei T, Zhou C H. 2012. Research progresses of surface temperature characteristic change over Tibetan Plateau since 1960. Progress in Geography, 31(11): 1503-1509.]
- 孙林岩. 1990. 组合预测方法及应用. 预测, 9(4): 31-35. [Sun L Y. 1990. Combination forecasting method and its application. Forecasting, 9(4): 31-35.]
- 汤国安, 杨昕. 2006. ArcGIS地理信息系统空间分析教程. 北京: 科学出版社. [Tang G A, Yang X. 2006. Tutorial of ArcGIS spatial analysis. Beijing, China: Science Press.]
- 汤青, 徐勇, 李扬. 2013. 黄土高原农户可持续生计评估及未来生计策略: 基于陕西延安市和宁夏固原市1076户农户调查. 地理科学进展, 32(2): 161-169. [Tang Q, Xu Y, Li Y. 2013. Assessment of farmers' sustainable livelihoods and future strategies on the Loess Plateau: based on a survey of 1076 farmers in Yan'an City in Shaanxi Province and Guyuan City in Ningxia Hui Autonomous Region. Progress in Geography, 32(2): 161-169.]
- 田广智, 杨生岳. 1981. 高原健康人55名呼气流速—容量曲线测定结果(摘要). 西北国防医学杂志, (3): 200-201. [Tian G Z, Yang S Y. 1981. Results in 55 healthy adults' expiratory flow volume curve (abstract). Medical Journal of National Defending Forces in Northwest China, (3): 200-201.]
- 王金玉, 李盛, 王式功, 等. 2013. 沙尘污染对暴露人群呼吸系统健康的影响. 中国沙漠, 33(3): 826-831. [Wang J Y, Li S, Wang S G, et al. 2013. Effects of dust pollution on respiratory system of long term exposed population. Journal of Desert Research, 33(3): 826-831.]
- 王涛, 范勇, 彭宝珠, 等. 2000. 高原环境对人肺功能的影响. 医学动物防制, 16(7): 381-383. [Wang T, Fan Y, Peng B Z, et al. 2000. Effect of oxygen shortage on blood pressure in plateau environment. Chinese Journal of Pest Control, 16(7): 381-383.]
- 王瑶, 张汉丰, 关玉珍, 等. 1988. 呼市地区健康成人最大呼气流速—容量曲线测定. 内蒙古医学杂志, 8(3): 154-155. [Wang Y, Zhang H F, Guan Y Z, et al. 1988. Determination of Hohhot healthy adults' maximal expiratory flow volume curve. Inner Mongolia Medical Journal, 8(3): 154-155.]
- 王兆维, 宋珠元, 文金序, 等. 1984. 广州地区健康人1011例肺通气功能的检测. 第一军医大学学报, 4(4): 296-299. [Wang Z W, Song Z Y, Wen J X, et al. 1984. Detection of 1011 healthy people in Guangzhou of pulmonary ventilation function. Journal of First Military Medical University, 4(4): 296-299.]
- 夏国恩, 金炜东, 张葛祥. 2006. 基于支持向量分类机和回归机的综合评价方法. 西南交通大学学报, 41(4): 522-527. [Xia G E, Jin W D, Zhang G X. 2006. Synthetic evaluation method based support vector classifier and re-

- gression machine. *Journal of Southwest Jiaotong University*, 41(4): 522-527.]
- 谢美云, 武希强, 薄晓霞, 等. 2012. 蒙古族正常人群呼气峰流速的测定. *包头医学院学报*, 28(3): 16-18. [Xie M Y, Wu X Q, Bo X X, et al. 2012. Determination of peak expiratory flow rate in normal Mongolian population. *Journal of Baotou Medical College*, 28(3): 16-18.]
- 徐风珍, 刘玉珠. 1985. 134名健康成人最大呼气流速—容量曲线的测定. *蚌埠医学院学报*, 10(2): 114-117. [Xu F Z, Liu Y Z. 1985. Measurements of maximal expiratory flow- volume curves in 134 normal adults. *Journal of Bengbu Medical College*, 10(2): 114-117.]
- 徐建华. 2006. 计量地理学. 北京: 高等教育出版社. [Xu J H. 2006. *Measurement geography*. Beijing, China: Higher Education Press.]
- 徐芝英, 胡云锋, 刘越, 等. 2012. 空间尺度转换数据精度评价的准则和方法. *地理科学进展*, 31(12): 1574-1582. [Xu Z Y, Hu Y F, Liu Y, et al. 2012. A review on the accuracy analysis of spatial scaling data. *Progress in Geography*, 31(12): 1574-1582.]
- 阎崇年, 颜吉鹤, 宋俊岭, 等. 1991. 中国市县大辞典. 北京: 中共中央党校出版社. [Yan C N, Yan J H, Song J L, et al. 1991. *Dictionary of Chinese cities and counties*. Beijing, China: The CPC Central Party School Press.]
- 颜宏, 沈国权, 毛耀顺, 等. 2002. 中华人民共和国气候图集. 北京: 气象出版社. [Yan H, Shen G Q, Mao Y S, et al. 2002. *The People's Republic of China climatic maps*. Beijing, China: Meteorological Press.]
- 杨楠. 2004. 岭回归分析在解决多重共线性问题中的独特作用. *统计与决策*, (3): 14-15. [Yang N. 2004. Unique function of ridge regression in solving multicollinearity problems. *Statistics and Decision*, (3): 14-15.]
- 杨青生, 张红贤, 葛森, 等. 2006. 基于人工神经网络的老年男性血沉参考值与地理环境关系研究. *地理科学*, 26(6): 749-754. [Yang Q S, Zhang H X, Ge M, et al. 2006. Relationship between reference value (Wintrobe) of Chinese old men's erythrocyte sedimentation rate (ESR) and geographical factors based on NN. *Scientia Geographica Sinica*, 26(6): 749-754.]
- 杨生岳, 田广智, 赵永吉. 1991. 不同海拔高度健康人506例最大呼气流量—容积曲线的研究. *西北国防医学杂志*, 12(1): 24-25. [Yang S Y, Tian G Z, Zhao Y J. 1991. Study on different altitude of 506 healthy people's maximal expiratory flow volume curve. *Medical Journal of National Defending Forces in Northwest China*, 12(1): 24-25.]
- 张晓明, 裴秀坤, 王勇, 等. 1995. 空气污染与儿童呼出气最大流速的研究. *环境与健康杂志*, 12(2): 53-55. [Zhang X M, Pei X K, Wang Y, et al. 1995. Study on air pollution and peak expiratory flow. *Journal of Environment and Health*, 12(2): 53-55.]
- 赵海燕, 陈宝元, 郭美南. 2003. 夜间哮喘呼气流速峰值及呼吸驱动昼夜节律的研究. *天津医科大学学报*, 9(1): 27-28, 32. [Zhao H Y, Chen B Y, Guo M N. 2003. Study on circadian rhythm of PEF and respiratory drive in nocturnal asthma. *Journal of Tianjin Medical University*, 9(1): 27-28, 32.]
- 赵济, 陈传康, 伍光和, 等. 1999. 中国地理. 北京: 高等教育出版社. [Zhao J, Chen C K, Wu G H, et al. 1999. *Geography of China*. Beijing, China: Higher Education Press.]
- 赵立, 姜洪芳, 朱丽华, 等. 2009. 沈阳地区健康成人肺功能与国外预计值比较分析. *中国实用内科杂志*, 29(9): 830-832. [Zhao L, Jiang H F, Zhu L H, et al. 2009. The difference between oversea predicted values and Chinese healthy adult measurements of spirometry in Shenyang. *Chinese Journal of Practical Internal Medicine*, 29(9): 830-832.]
- 赵蓉雅. 2010. 上海地区肺功能正常预计值公式的时效性检验及肺功能医学参考值范围的初步研究[D]. 上海: 复旦大学. [Zhao R Y. 2010. *The chronergy of prediction equations and preliminary study of medical reference range for normal lung function in Shanghai*[D]. Shanghai, China: Fudan University.]
- 赵瑛. 1999. 人体昼夜节律和健康. *生物学通报*, 34(8): 18-20. [Zhao Y. 1999. The body's circadian rhythms and health. *Bulletin of Biology*, 34(8): 18-20.]
- Faruto. 2010. LIBSVM- faruto ultimate version: a toolbox with imple-ments for support vector machines based on libsvm[EB/OL]. April 18, 2010 [October 5, 2013]. <http://www.ilovematlab.cn/forum.php?mod=viewthread&tid=74019>
- Getis A, Ord J K. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3): 189-206.
- Leuenberger P, Kunzli N, Ackermann LU, et al. 1998. Swiss study on air pollution and lung diseases in adults. *Schweiz Med Wochenshr*, 12(8): 105-161.

Application of combination forecasting model in geographic distribution of reference value of women's peak expiratory flow rate

XUE Ranyin, GE Miao, HE Jinwei, HU Yanyu, GU Linlin, YANG Shaofang

(Health Geography Institute of Tourism and Environment College, Shaanxi Normal University, Xi'an 710119, China)

Abstract: With the development of geography and people's overall health concerns, medical geography as an emerging discipline has also experienced rapid developments. Given that most of the existing medical reference values take into little consideration the influence of geographical factors, a more comprehensive and scientific method should be developed to take these into account. This article takes healthy adult women's lung peak expiratory flow rate reference value as an example, using 3809 cases of healthy adult women's peak expiratory flow reference value collected throughout China to analyze the impact of geographic factors, calculate the differences of different regions' reference values, and explore the mechanism of geographical factor's influence on medical reference value in an effort to improve the methods of medical reference value analysis through analyzing the relationship between geographical factors and medical reference values. As a first step, correlation analysis was used to analyze the relationship between peak expiratory flow value of adult women and the selected 25 indicators of geographical factors. Based on the result, seven geographic indicators (latitude, altitude, average temperature, annual average relative humidity, annual rainfall, topsoil gravel percentage, and topsoil reference capacity) that have significant correlation with peak expiratory flow reference value were extracted for further analysis. Second, Moran's I (spatial autocorrelation module), one of the ArcGIS software's analytical tools, was used to determine if this group of data is impacted by spatial and geographical factors. Third, using the data for the seven selected indicators, ridge regression analysis and SVR (support vector regression) were used to create two regression models and interpolate values. Then the results of these two prediction models were given different weights to establish the optimal combination forecasting model of spatial differences. Student's T test was used to compare the accuracy of ridge regression analysis, SVR and the combination forecasting model. Meanwhile, differences between the true values and results of the above three models were also considered for evaluating the performance of the models. Finally, a spatial difference prediction map was made. Based on this map and the results of correlation analyses, this article discusses why and how these geographical factors influence human tissues/organs and medical reference values. The output of this study indicates that the relationship between geographical factors and healthy adult Chinese women's lung peak expiratory flow rate should not be overlooked. The selected geographical factors (classify into terrain, climatic and soil factors) affect the lung tissue, especially the structure and function of the bronchi, because different living environments impact human tissues and organs differently, and humans living in different regions develop some differences in tissues and organs. The result of this research also shows that the combination forecasting model, which combined ridge regression and SVR, performed better than the individual prediction methods. Combination forecasting model not only can be used in traditional prediction exercises using temporal data: it is also possible to use this method to predict differences in geographic distribution or spatial data. It has the potential to be further expanded and utilized.

Key words: peak expiratory flow rate; ridge regression; support vector regression; combination forecasting model; geographical distribution; China