

基于地统计学插值方法的洪湖水质时空状况分析

吴红艳¹ 程东升² 周文兵¹ 胡晓波¹ 葛晓东¹ 朱端卫¹

1. 华中农业大学资源与环境学院, 武汉 430070; 2. 中国水科院水环境研究所, 北京 100038

摘要 以洪湖 2007—2010 年间水质因子为研究对象, 利用地统计学和常用的 5 种插值方法(径向基函数、反距离权重、全局多项式、局部多项式和普通克里格插值法)对洪湖水质的时空状况进行分析, 并用误差评价因子(MAE、MRE 和 RMSE)对其拟合情况进行评价。结果表明, 各水质因子均具有较高的时空异质性, 其中, 总氮质量浓度受到温度、水深以及水循环的影响; 在 5 种插值方法中, 径向基函数取得了较好的模拟预测结果, 洪湖总氮质量浓度具有不同的空间分布特征。

关键词 水质因子; 地统计学; 时空特征; 插值模型; 误差评价因子; 洪湖

中图分类号 P 333.1 **文献标识码** A **文章编号** 1000-2421(2013)03-0082-06

洪湖是长江和汉水支流东荆河之间的大型浅水洼地壅塞湖, 是湖北省最大的淡水湖泊。洪湖还是我国主要的淡水鱼类生产基地、淡水供应地以及农业蓄水排涝调节地区, 同时也是航运和旅游的主要载体^[1]。近年来由于不合理的利用和开发, 洪湖的生态环境发生了较大的变化, 尤其是水环境状况^[2]。了解洪湖水质状况的改变情况, 并进行相应的分析, 可以更好地保护洪湖湿地的生态环境。由于洪湖水水质参数呈现复杂的空间分布特征和空间非平稳性, 且易受到环境因子的影响。因此, 根据监测数据选择合适的模型构建准确的洪湖水水质预测模型, 对指导洪湖的生态保护将起到重要作用。

地统计学是以区域化变量为基础, 借助变异函数, 使得研究既具有随机性又具有结构性的一门学科^[3-4]。刘瑞民等^[5]基于不同尺度的太湖水质参数, 探讨了其空间分布和结构异向性, 得出污染物集中在湖岸线附近, 不易向湖中央扩散, 各水质参数空间分布差异性较大。卢进登等^[6]基于 GIS 技术对长湖水污染状况进行研究, 得出了污染物的主要空间分布和长湖水质的分类。刘瑞民等^[7]基于克里格插值模型对太湖水质参数进行了研究, 并对其污染水平进行了分类。上述研究虽是地统计学在湖水生态环境研究中的应用, 但研究者采用的方法大多只局限于一种, 没能提供多种参考模型以证明地统计学

模型的优劣。地统计学模型在水质参数空间应用上容易受到外界环境和时间的影响, 模型不仅具有空间变异性还具有时间变异性。另外, 在考虑水质参数空间结构分布特征和时间变异的同时还需要考虑其局部变异性, 以更好地研究水质参数的空间非平稳性和空间依赖性。为此, 笔者选取在地统计学中常用的全局多项式、局部多项式、径向基函数、反距离权重和普通克里格插值法等 5 种模型, 从不同的方面探讨洪湖水水质参数的时空特征, 以此得出合适的模型来预测和模拟洪湖水水质的变化情况, 从而为保护洪湖的生态环境提供一定的科学依据和理论支持。

1 材料与方法

1.1 研究区域概况

洪湖位于长江中游江汉湖群, 处在四湖(长湖、三湖、白露湖、洪湖)流域最下游, 地处湖北省洪湖市西南部, 跨洪湖、监利两市县, 北纬 $29^{\circ}42' \sim 29^{\circ}58'$, 东经 $113^{\circ}13' \sim 113^{\circ}29'$, 东西长 23.4 km, 南北宽 20.8 km, 面积为 348.2 km², 湖底自西向东略有倾斜, 西浅东深。北亚热带湿润季风气候是此区的基本气候特征, 表现出热能充足、降水丰富、雨热同季的气候特点^[8]。

1.2 数据来源

1) 样品采集及测量指标。根据洪湖的实际地理

收稿日期: 2012-05-20

基金项目: 国家科技重大水专项项目(2009ZX07631)

吴红艳, 硕士研究生, 研究方向: 水污染防治政策的社会经济影响评价, E-mail: wuhongyan@webmail.hzau.edu.cn

通讯作者: 周文兵, 博士, 副教授, 研究方向: 农业资源利用, E-mail: zhouwb@mail.hzau.edu.cn

位置和几何形状,同时考虑洪湖的污染状况和出入水口等地理位置,自 2007 年在杨柴湖、桐梓湖、蓝田、湖心 A、湖心 B、小港、下新河和排水闸等处设置 8 个监测点,每月在各样点处用取水器沉至水下 30~50 cm 采集水样,每样点采集 3 次,水样用于水温、水深、pH 值、透明度(SD)、化学需氧量(高锰酸盐指数,COD_{Mn})、溶解氧(DO)、氨氮(AN)、五日生化需氧量(BOD₅)、总磷(TP)、总氮(TN)、叶绿素 a 等水质因子的测定。其中水温、水深、透明度,在采集水样时现场测定,其余各因子在水样带回实验室后的 1 周内测定完毕。

2)测量方法。氨氮用纳氏试剂比色法测定,化学耗氧量采用高锰酸盐指数酸性法,总氮采用硫酸钾氧化—紫外分光光度法测定,总磷采用钼锑抗分光光度法,叶绿素过滤后用分光光度法测定,五日生化需氧量采用微生物电极法测定^[9]。

面更依赖于局部数据的变异。(5)径向基函数插值法(radial basis function,RBF),是多个数据插值方法的组合,即经过各个已知样点生成一个圆滑曲面,并使表面的总曲率最小^[12]。

1.4 插值精度评价

采用交叉验证法^[13]评价各种方法的插值效果,将每个点作为预测点,其余点则作为建模点,分析预测值和真实值之间的误差。误差评价指标采用平均绝对误差(MAE)、平均相对误差(MRE)和均方根误差(RMSE)。其中,MAE 反映估计值的误差范围,MRE 反映估计值对于观测值的准确度,RMSE 则反映估计值的灵敏度和极值情况^[14-15],表达式为:

$$MAE=\frac{\sum_{i=1}^n|Z_{a,i}-Z_{c,i}|}{n} \tag{1}$$

$$MRE=\frac{1}{n}\sum_{i=1}^n\frac{|Z_{a,i}-Z_{c,i}|}{Z_{a,i}} \tag{2}$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^n(Z_{a,i}-Z_{c,i})^2} \tag{3}$$

其中: $Z_{a,i}$ 、 $Z_{c,i}$ 分别为第 i 个站点的实际观测值和插值预测值, n 为验证站点数。分别从时间和空间 2 个角度来探讨洪湖水质因子状况,先以年为单位分析水质因子随时间的变化情况,然后分别分析 8 个采样点水质因子的空间相关性,以及水质因子在时间和空间上的异质性,并通过误差评价指标得出最佳的预测拟合模型。

2 结果与分析

2.1 水质因子时间特征分析

共获取 2007—2010 年每月的水质检测报告,提取了 11 个水质因子:水深、水温、透明度、pH 值、溶解氧(DO)、化学需氧量(高锰酸盐指数)、五日生化需氧量(BOD₅)、氨氮(AN)、总氮(TN)、总磷(TP)和叶绿素 a,共 $4\times12\times11=528$ 个研究因子,分析每年中水质因子平均值的基本特征(表 1)。

表 1 表明,水质因子参数随着时间的变化整体变化幅度不大,除水温、pH 值和叶绿素 a 之外,其他变化幅度均小于 1,特别是总磷和氨氮,在 4 a 中的变化幅度最小,并且除水温的标准差和方差大于 1 外,其余的均小于 1,说明水质因子与均值的偏离程度很小,即水质因子的年均值随时间的变化情况不大。因此,以月为时间单位分析各点的水质因子状况,同时为消除异常值和极值的影响,求取 4 a 中各点各月的平均值,得出基本信息特征(表 2)。

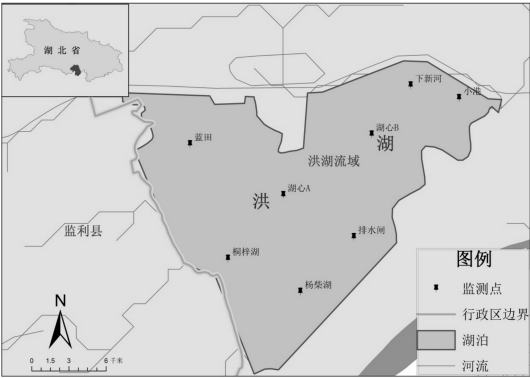


图 1 洪湖常规监测点位图

Fig. 1 Routine monitoring point location of Honghu Lake

1.3 研究方法

空间插值是利用已知的部分空间样本信息对未知的地理空间特征进行估计,也是 GIS 的重要功能模块之一^[10]。Dirks 等^[11]将空间插值方法又分为整体插值法和局部插值法。利用常用的 5 种插值方法:(1)反距离权重插值法(inverse distance weighted,IDW),以插值点与样本点间的距离为权重进行加权平均,离插值点越近的样本点赋予的权重越大。(2)普通克里格插值法(ordinary Kriging,OK),是区域化变量的线性估计,但其考虑了空间相关性,更加符合空间数据的特点。(3)全局多项式插值法(global polynomial,GP),以整个研究区的样点数据集为基础,用一个多项式计算预测值。(4)局部多项式插值法(local polynomial,LP),采用处在特定重叠邻近区域内的多个多项式进行插值,其产生的曲

表 1 洪湖水质因子年度基本信息¹⁾

Table 1 Annual basic information of water quality factors of Honghu Lake

项目 Item	D/m	T/℃	Tr/m	pH	DO/ (mg/L)	COD _{Mn} / (mg/L)	BOD ₅ / (mg/L)	AN/ (mg/L)	TN/ (mg/L)	TP/ (mg/L)	Cha/ (mg/L)
2007	1.56	18.98	0.51	7.04	9.37	4.25	2.88	0.33	1.21	0.04	3.47
2008	1.43	20.61	0.93	6.89	9.15	4.43	3.00	0.38	1.01	0.04	2.88
2009	1.75	18.50	1.04	8.01	9.85	4.62	3.27	0.27	0.75	0.02	3.49
2010	1.86	18.33	1.31	7.30	9.39	3.75	2.70	0.34	0.77	0.04	3.97
变化范围 Variation range	0.43	2.28	0.80	1.11	0.69	0.87	0.57	0.11	0.46	0.02	1.09
极小值 Minimum	1.43	18.33	0.51	6.89	9.15	3.75	2.70	0.27	0.75	0.02	2.88
极大值 Maximum	1.86	20.61	1.31	8.01	9.85	4.62	3.27	0.38	1.21	0.04	3.97
均值 Mean	1.65	19.10	0.95	7.31	9.44	4.26	2.96	0.33	0.93	0.04	3.45
标准差 Standard deviation	0.19	1.04	0.33	0.49	0.29	0.37	0.24	0.04	0.22	0.01	0.45
方差 Variance	0.037	1.080	0.111	0.244	0.090	0.139	0.058	0.002	0.047	0.000	0.190

1)D:水深 Depth; T:水温 Temperature; Tr:透明度 Transparency; DO:溶解氧 Dissolved oxygen; COD_{Mn}:化学需氧量 Chemical oxygen demand; BOD₅:五日生化需氧量 Biochemical oxygen demand after 5 days; AN:氨氮 Ammonia nitrogen; TN:总氮 Total nitrogen; TP:总磷 Total phosphorus; Cha:叶绿素 a Chlorophyll a; 下同 The same as below.

表 2 不同监测点水质因子基本信息特征

Table 2 Basic information of water quality factors from different monitoring stations

项目 Item	D/m	T/℃	Tr/m	pH	DO/ (mg/L)	COD _{Mn} / (mg/L)	BOD ₅ / (mg/L)	AN/ (mg/L)	TN/ (mg/L)	TP/ (mg/L)	Cha/ (mg/L)
全距 Range	1.450	24.625	1.075	1.050	6.700	2.410	4.000	0.475	1.325	0.061	5.760
极小值 Minimum	1.150	4.750	0.575	6.678	6.650	2.950	0.800	0.088	0.443	0.013	0.853
极大值 Maximum	2.600	29.375	1.650	7.728	13.350	5.360	4.800	0.563	1.768	0.074	6.613
总和 Total	156.382	1 834.075	90.457	701.756	913.638	411.088	278.100	29.331	85.834	3.277	333.394
均值 Mean	1.629	19.104	0.942	7.310	9.517	4.282	2.897	0.306	0.894	0.034	3.473
标准差 Standard deviation	0.314	8.128	0.211	0.265	1.557	0.454	0.826	0.095	0.213	0.014	1.242
方差 Variance	0.098	66.070	0.044	0.070	2.423	0.206	0.683	0.009	0.045	0.000	1.542
变异系数 Variation coefficient	0.193	0.426	0.225	0.036	0.164	0.106	0.285	0.312	0.238	0.404	0.358
P	0.197	0.000	0.943	0.007	0.961	0.920	0.606	0.646	0.158	0.140	0.826

由于普通克里格方法要求插值样本最好呈正态分布,为此首先对洪湖水质参数进行统计分析,并通过 K-S 检验来验证数据的正态分布情况,通过 K-S 检验结果 P 值可知,除水温外其余水质参数均符合正态分布检验结果($P>0.05$),可以采用地统计学方法对其进行插值处理。同时,从表 2 可知,总磷、水温和叶绿素 a 的变异系数最大,分别达到了 40.4%、42.6%和 35.8%,说明这些水质因子在时空上存在一定的变异性。对照《地表水环境质量标准》(GB3838—2002)可以发现,不同的水质因子跨越不同的标准等级,其中,总氮的跨度最大,跨越了 4 级,其余水质因子均跨越 2~3 级。

下面以总氮为例进行分析。图 2 为各观测点总氮质量浓度随时间的变化情况,从中可以发现,总氮质量浓度随时间变化表现出不同的特征,而不同点的总氮质量浓度在同一时间上则表现出相似的规律

性。综合比较发现,总氮质量浓度在 3 月和 11 月时有较大值,而 5 月和 7 月则有较小值,在 10 月至次年 2 月一直处于较大值。根据洪湖的实际情况可知,当温度较高时总氮质量浓度会有较小的值,而当温度较低时则有较大的值,这主要是由于总氮为硝酸盐氮、亚硝酸盐氮、氨氮和有机氮的总称,是反映水体富营养化的主要指标,当温度升高时水中微生物大量繁殖,水体硝化与反硝化促进氮损失,且浮游生物生长旺盛会消耗大量的氮,而温度降低时水体中的生物量减少,总氮累积,浓度增大。因此,总氮不仅可以作为衡量水质的重要指标之一,同时也可以间接反映水中微生物和浮游生物的旺盛程度。由图 2 可知,蓝田监测点的总氮质量浓度相对较高,而排水闸和小港附近的总氮质量浓度较低,这主要是由于水流对湖水中总氮质量浓度的影响。总之,可根据不同监测点的总氮情况提出有针对性的措施,

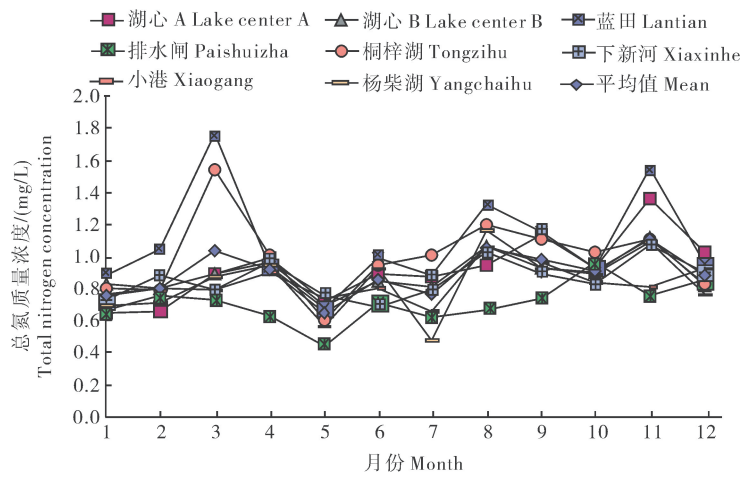


图 2 洪湖各监测点总氮质量浓度在一年中的变化情况

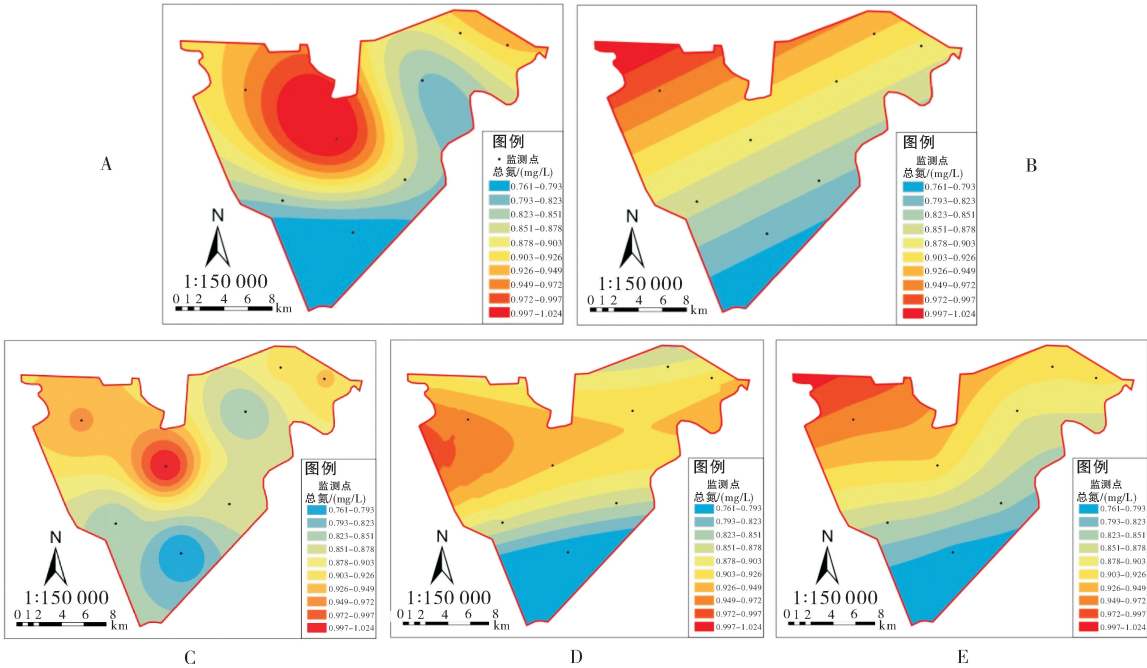
Fig. 2 The change of total nitrogen concentration in a whole year of Honghu Lake

对洪湖总氮质量浓度过高的区域进行有效治理,以有效降低人力和物力成本。

2.2 水质因子空间特征分析

虽然不同插值方法对洪湖总氮质量浓度的模拟预测得到不同的模拟图(图 3),但整体上表现出了一致性,总氮质量浓度较高的区域一般集中在洪湖的西北部(湖心 A 和蓝田区域),而浓度较低的区域则主要分布在南部(杨柴胡和排水闸区域)。洪湖南

部和东部相对于西北部来说水较深,并且有排水闸和小港可以使得湖水形成循环,不易囤积大量的废物,且水深则容易自我净化,因此总氮质量浓度较低。比较不同插值方法可以发现,径向基函数法和反距离权重法以观测点数据为主,呈现出圆环状的分布;全局多项式法和局部多项式法拟合结果图形则相对较平滑,主要考虑的全局因素;而克里格插值法则介于两者之间,在不同的地点形成不同的拟合图。



A: 径向基函数法 Radial basis function; B: 全局多项式法 Global polynomial; C: 反距离权重法 Inverse distance weighted; D: 普通克里格法 Ordinary Kriging; E: 局部多项式法 Local polynomial.

图 3 不同插值方法对洪湖总氮质量浓度的模拟预测图

Fig. 3 The forecast draft of total nitrogen concentration of different interpolation method

根据误差评价指标得出 5 种插值方法的评价结果(表 3)。整体比较发现,5 种模型均表现出较好的预测结果,并且 MAE、MRE 和 RMSE 的值均小于 0.11,有较好的拟合度,通过交叉验证得到的预测结果与实际值相接近;比较发现,径向基函数法<全局多项式法<反距离权重法<普通克里格法<局部多项式法。由于 5 种插值方法所采用的模型不同,因此所遵循的理论依据也不同,单纯地从误差评价指标中可以看出,径向基函数法对 2010 年 12 月的洪湖水水质指标总氮预测模拟具有较好的拟合度。

表 3 不同插值方法对洪湖总氮质量浓度模拟的误差评价指标值¹⁾

Table 3 The error evaluation index value of different interpolation methods used to forecast the concentration of total nitrogen of Honghu Lake			
项目 Item	MAE	MRE	RMSE
径向基函数法 Radial basis function	0.061 491	0.069 406	0.086 875
全局多项式法 Global polynomial	0.072 346	0.081 180	0.088 219
反距离权重法 Inverse distance weighted	0.072 824	0.082 604	0.095 237
普通克里格法 Ordinary Kriging	0.077 100	0.087 601	0.095 109
局部多项式法 Local polynomial	0.081 836	0.090 927	0.100 731

1)MAE:平均绝对误差 Mean absolute error; MRE:平均相对误差 Mean relative error; RMSE:均方根误差 Root mean square error.

3 讨 论

由于湖水水质具有明显的空间变异性和不稳定性,因此在时间和空间上的分布具有明显的不一致性,为此选取合适的模型并通过对一定数量的采样点水质的研究来估测全局湖水水质的分布特征十分必要。目前国内外学者通过不同的地统计模型对不同地区湖水的水质进行了研究,并通过比较不同精度的验证指标得出了适合的模型与适用条件。同时,基于不同的模型方法也可以探讨引起湖水水质变化的影响因子,并可以为分析湖水状况提供又一理论依据^[16-21]。王艳红等^[21]利用克里格内插值估计对太湖春季 15 项水质参数进行了估计,直观说明了太湖的污染状况;卢进登等^[22]利用反距离权重法和克里格方法对长湖水水质进行了分析,并比较得出克里格方法略优于反距离权重法;与本研究利用 5 种常用的插值方法对 2010 年 12 月总氮质量浓度的

模拟预测结果具有一定的相似性,其中,洪湖西北部的总氮质量浓度较高,而南部和东部较低,受到洪湖水自我净化以及水循环影响,总氮不容易发生集聚;同时根据误差评价标准,对 2010 年 12 月总氮质量浓度预测模拟后发现,径向基函数法在 MAE、MRE 或者 RMSE 上均具有一定的优势,这为洪湖水质的模拟预测的应用提供了一定的理论依据和支持。限于试验条件,本文仅获得了洪湖 8 个监测点的水质数据,且仅对 2010 年 12 月总氮质量浓度进行了预测和模拟,这对于整个洪湖水质的预测来说具有一定的局限性,在后续工作需要加强相关研究,为湖泊环境生态保护提供更多的理论支持。

参 考 文 献

[1] 蔡述明,杜耘. 江汉湖群湖泊资源特点及其开发利用保护[J]. 华中师范大学学报,2000,34(4):476-481.

[2] 姚桂枝,金卫兵,刘章勇,等. 洪湖水环境质量模糊综合评价[J]. 湖北农业科学,2009,48(6):1358-1360.

[3] 李晓晖,袁峰,贾蔡,等. 基于地统计学插值方法的局部奇异性指数计算比较研究[J]. 地理科学,2011,32(2):136-142.

[4] 夏斌,伍第政. 基于地统计学的广东省土壤铬空间变异分析[J]. 湖北农业科学,2011,50(21):4368-4370.

[5] 刘瑞民,王学军,郑一,等. 太湖水水质参数中小尺度空间结构特征[J]. 长江流域资源与环境,2002,11(1):32-35.

[6] 卢进登,张劲,陈红兵,等. GIS 技术在长湖水污染现状分析中的应用[J]. 湖北大学学报:自然科学版,2007,29(2):199-202.

[7] 刘瑞民,王学军,郑一,等. 地统计学在太湖水水质研究中的应用[J]. 环境科学学报,2002,22(2):209-212.

[8] 卢山,李世杰,王学雷. 洪湖的环境变迁与生态保护[J]. 湿地科学,2004,2(3):234-237.

[9] 奚旦立,孙欲生,刘秀英. 环境监测[M]. 北京:高等教育出版社,1995:378-393.

[10] TIAN Y, XIAO C C, LIU Y, et al. Effects of raster resolution on landslide susceptibility mapping: a case study of Shenzhen [J]. Science in China Series E: Technological Sciences, 2008, 51 (S2):188-198.

[11] DIRKS K N, HAY J E, STOW C D, et al. High resolution studies of rainfall on Norfolk Island: Part II. interpolation of rainfall data[J]. Journal of Hydrology, 1998, 208 (3/4): 187-193.

[12] 王树良,史文中,李德毅,等. 基于张力样条插值函数的土地数据挖掘[J]. 计算机工程与应用,2003,39(25):5-7.

[13] SEAMAN R S. Objective analysis accuracies of statistical in terpolation and successive correction schemes[J]. Australian Meteorological Magazine, 1983, 31: 225-240.

[14] 宋丽琼. 日降水量的空间插值方法与应用对比分析——以深圳市为例[J]. 地球信息科学,2008,10(5):566-572.

[15] 高华喜,殷坤龙. 降雨与滑坡灾害相关性分析及预警预报阈值之探讨[J]. 岩土力学,2007,28(5):1055-1060.

[16] 张春媛,贾克力,童伟. 普通克里格法在乌梁素海水体富营养化研究中的应用[J]. 北方环境,2011(7):100-102.

[17] 文航,蔡佳亮,苏玉,等. 利用水生生物指标识别滇池流域入湖河流水质污染因子及其空间分布特征[J]. 环境科学学报,2011,31(1):69-80.

[18] 武周虎,张可,金玲仁,等. 南四湖水质空间分布特征分析与改善效果评估[J]. 水资源保护,2012,28(6):1-7.

[19] 黎洁,单保庆,宋芬,等. 永定河和滦河水系浮游动物多样性调查与分析[J]. 华中农业大学学报,2011,30(6):768-774.

[20] 宋芬,王卫民,单保庆,等. 马颊河与徒骇河浮游植物群落特征及水质初步评价[J]. 华中农业大学学报,2011,30(3):364-370.

[21] 王艳红,张平,邓正栋. 太湖春季水质参数空间格局分布的克里格内插估计[J]. 解放军理工大学学报:自然科学版,2006,7(2):175-180.

[22] 卢进登,帅方敏,梁雄兵,等. 空间插值法在湖泊水污染现状评价中的应用研究[J]. 湖南科技大学学报:自然科学版,2007,22(3):125-128.

Temporal and spatial analysis of water quality of Honghu Lake
based on statistical interpolation methods

WU Hong-yan¹ CHENG Dong-sheng² ZHOU Wen-bing¹
HU Xiao-bo¹ GE Xiao-dong¹ ZHU Duan-wei¹

1. College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China;
2. Department of Water Environment Research, China Institute of Water Resource
and Hydropower Research, Beijing 100038, China

Abstract The water quality of Honghu Lake between 2007 and 2010 was taken as an example and the geostatistics and five commonly used interpolation methods (radial basis function, inverse distance weighted, global polynomial, local polynomial and ordinary Kriging) were used to analyze the temporal and spatial conditions of water quality of Honghu Lake. Error evaluation factors (MAE, MRE and RMSE) were employed to evaluate the fitting situation. It concluded that all factors of the water quality had high spatial-temporal heterogeneity and the concentration of total nitrogen was influenced by temperature, water depth and water cycle. Among the five kinds of interpolation methods, the radial basis function had a better predict result. The total nitrogen concentration of Honghu Lake had different characteristics of spatial distribution.

Key words water quality factor; geostatistics; space-time characteristics; interpolation model; error evaluation factors; Honghu Lake

(责任编辑:陆文昌)