

新安江流域降雨量空间插值方法和参数筛选

何文英¹, 毕孟飞¹, 李泽利², 王玉秋²

(1. 黄山市环境监测站, 安徽 黄山 245000; 2. 南开大学 环境科学与工程学院, 天津 300071)

摘要: 选取新安江流域 48 个雨量站点 2000–2010 年的降雨量数据以及地理地形资料, 利用反距离权重 (IDW)、径向基函数 (RBF)、普通克里金 (OK) 和残差克里金 (RK) 方法进行插值, 根据交叉验证结果筛选最佳的方法和参数。结果表明: 年尺度上, 考虑高程、纬度等地形因子影响的 RK 方法插值精度最高, 然后是 IDW、OK 和 RBF; 邻近站点的搜索策略对 IDW 插值精度影响较大; 考虑各向异性的 OK 法插值精度较高, 模型之间的差别不大。月尺度上, IDW 法对降雨量丰富的月份插值精度较高; RBF 对降雨较少且分布均匀的月份插值效果好; 数据越符合正态分布, OK 法的插值精度越高; RK 方法则适用于降雨与地理因素相关性好的月份。

关键词: 空间插值; 反距离权重; 径向基函数; 克里金方法; 新安江

中图分类号: P332.1

文献标识码: A

文章编号: 1672-643X(2013)01-0130-05

Spatial interpolation method and parameters screening for precipitation in Xin'anjiang river basin

HE Wenying¹, BI Mengfei¹, LI Zeli², WANG Yuqiu²

(1. Environment Monitoring Station of Huangshan City, Huangshan 245000, China;

2. College of Environmental Science and Engineering, Nankai University, Tianjin 300071, China)

Abstract: The paper selected precipitation and geographic terrain data of 48 rainfall gage stations in the Xin'anjiang River basin from 2000 to 2010, and used inverse distance weight (IDW), radial basis function (RBF), ordinary Kriging (OK) and residual Kriging (RK) methods to interpolate and screened the best interpolation method and parameter according to the validated results. The results show that in annual scale the precision of RK method with comprehensively considered the influences of elevation and latitude is the highest, then follows by IDW, OK and RBF; the search strategy of neighbor stations has greater impact on IDW interpolation accuracy; the precision of OK method with considered anisotropic is higher, but the difference between the models is little. On monthly scale, the interpolation precision of IDW method to rich rainfall month is higher; the interpolation effect of RBF method to less and evenly distributed rainfall month is good; the data is more in accordance with normal distribution, the higher the interpolation accuracy of OK method; RK method is applicable to the month when the correlation between rainfall distribution and geographical factors is good.

Key words: spatial interpolation; inverse distance weight; radial basis function; Kriging method; Xin'anjiang River

降雨时空分布的不均匀性对流域产汇流的影响非常大^[1]。由于研究区地形地貌、技术手段等多方面因素的综合限制, 绝大多数空间位置上的数据是无法获得的。根据已有站点的监测数据, 选取合理而准确的空间插值技术获得连续的降水曲面一直是学者所研究的热点, 如: 不同插值技术精度的比

较^[2-5]、插值技术内部参数的筛选^[2,6]、对传统插值技术的改进和修正^[7-9]以及非传统技术的开发应用^[1,10]等。然而研究发现并没有一种普适的插值方法和固定的参数取值是最优的, 同时由于研究区域特点和数据集的不同, 插值方法的选择和内部参数的筛选显得十分必要。本文基于新安江流域的降雨和地形资料, 利用反距离权重 (IDW)、径向基函数

收稿日期: 2012-09-25; 修回日期: 2012-10-11

基金项目: 环保部环境规划院水污染防治综合项目 (2012A012)

作者简介: 何文英 (1966-), 男, 安徽安庆人, 硕士研究生, 高级工程师, 主要从事环境监测工作。

通讯作者: 王玉秋 (1965-), 男, 黑龙江齐齐哈尔人, 教授, 博士生导师, 主要从事水环境管理决策与技术支持, 流域及水源地保护战略等研究。

(RBF)、普通克里金(OK),以及残差克里金(RK)技术,通过对比不同插值方法和参数取值的结果,寻求最优的插值方案,为该流域水文、水质模型的应用提供可靠的降雨数据,对区域水资源管理和生态环境治理等都具有重要意义。

1 材料和方法

1.1 区域概况及资料整理

新安江发源于安徽省黄山市,注入浙江省千岛湖,流域地处中亚热带,属于湿润性季风气候,多年平均温度 15℃ ~ 16℃,降水丰富,年平均降雨量 1 700 mm 左右,但时空分布不均:时间上梅雨期降水集中,秋冬季降水较少,空间上年降水量大体上南多北少,受地形因素影响明显。研究使用美国太空总署和国防部国家测绘局联合测量的 DEM(图 1,分辨率 30 m × 30 m),经过 GIS 拼接,再通过流域边界裁剪后得到。

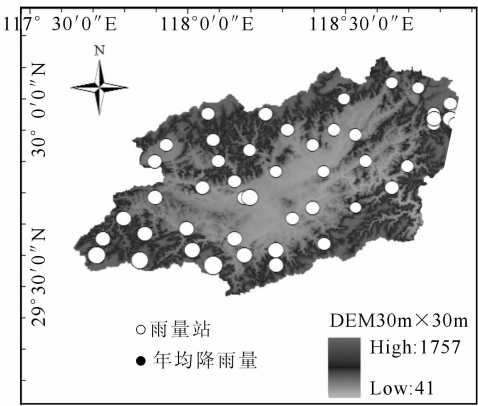


图 1 新安江流域地理位置,DEM 及雨量站分布图

为了研究流域雨量站之间降雨量的时空差异,对其 2000 - 2010 年的日降雨数据进行统计分析(图 1 和表 1)。表 1 中“均值”表示 48 个站点多年数据的月平均值及年平均值,可以看出降雨在时间分布上有较大差异,4 - 7 月份降水相对丰富;最大值和最小值表示基于月统计及年统计的 48 个站点中的最大值和最小值,说明了降雨在空间分布存在差异:图 1 给出了年均降雨量的空间分布,白色圆圈越大,代表该站点的年降雨量越多,可以看出,低纬度地区降水相对丰沛,同时流域海拔较高的区域(紫色区域)降水相对较大。

1.2 研究方法

研究基于 ArcGIS 9.3 的地统计分析模块,所采用的插值方法的基本原理及筛选的参数见表 2。由于每种插值方法内部都有诸多参数变量,在此仅对

插值结果影响较大的参数进行了研究。

表 1 雨量站的月尺度和年尺度统计结果 mm

时段	降雨量			
	均值	标准偏差	最小值	最大值
1 月	94.1	8.9	77.8	117.1
2 月	136.1	17.7	105.5	181.4
3 月	149.9	13.4	122.8	187.3
4 月	185.0	25.7	149.9	249.0
5 月	217.5	35.7	151.8	312.2
6 月	270.3	27.3	230.7	378.1
7 月	185.0	23.6	150.3	254.8
8 月	154.3	20.1	118.3	213.3
9 月	66.4	9.9	54.6	95.6
10 月	63.0	8.5	50.0	106.2
11 月	80.3	7.1	71.6	103.9
12 月	59.2	6.4	50.3	78.1
年均	1161.0	170.9	1383.9	2206.6

1.3 验证方法

采用全交叉验证的方法,利用地统计模块中导出的交叉验证结果,计算验证点的实测值与预测值之间的误差。评价插值误差大小的方法有很多种^[3],通常采取以下 4 种指标:

① 平均误差:

$$MBE = \frac{1}{n} \sum_{i=1}^n [\hat{Z}(x_i) - Z(x_i)]$$
, 此值越接近于零插值效果越好;

② 平均绝对误差:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{Z}(x_i) - Z(x_i)|$$
, 此值越小插值效果越好;

③ 均方根误差:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [\hat{Z}(x_i) - Z(x_i)]^2}$$
, 此值越小插值效果越好;

④ 线性相关系数:

$$CC = \frac{\sum_{i=1}^n [\hat{Z}(x_i) - \hat{Z}_m][Z(x_i) - Z_m]}{n \cdot \sigma_e \cdot \sigma_0}$$
, 此值越接近于 1 插值效果越好。

式中: $\hat{Z}(x_i)$ 和 $Z(x_i)$ 分别是第 i 个样本的预测值和观测值; $\hat{Z}_m$ 和 Z_m 分别为降雨量预测值和观测值的平均值; σ_e 和 σ_0 分别为降雨量估计值和观测值标准偏差; MBE 、 MAE 和 $RMSE$ 的单位是 mm, CC 无量纲。

表 2 插值方法的基本原理及参数选择

插值方法	公式	公式解释	筛选参数
IDW ^[5]	$Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i),$ $\lambda_i = D_i^{-P} / \sum_{i=1}^n D_i^{-P}$	$Z(x_0)$: x_0 点的估计值, $Z(x_i)$:第 i 个样本观测值, n :搜索样本的数量, λ_i 为距离权重, D_i : x_0 与 i 之间的欧式距离, P :幂指数。	幂指数 P 、搜索邻近站点数量 n 和搜索区域形状
RBF ^[4]	$Z(x) = \sum_{i=1}^n a_i \phi(d_i) + \sum_{j=1}^m b_j f_j(x)$	$\phi(d_i)$ 为径向基函数, d_i 为插值点 i 和采样点 x 之间的距离, $f_j(x)$ 为趋势函数。	3 种径向基函数:完全规则样条(CRS)、反转多重二次曲面(IMQ)、张力样条(ST)
OK ^[11]	$Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i),$ $\sum_{i=1}^n \lambda_i C(x_i, x_j) - \mu = C(x_i, x_0)$ 约束条件: $\sum_{i=1}^n \lambda_i = 1$	$Z(x_i)$ 为第 i 个样本观测值, n 为样本的数量, λ_i 为权重, $C(x_i, x_j)$ 为样本点之间的协方差, $C(x_i, x_0)$ 为样本点与插值点之间的协方差, μ 为拉格朗日乘子。	3 种理论半方差模型:球面(SPH)、指数(EXP)和高斯(GS),同时比较各向同性和各向异性 ^[6] 的插值结果。
RK ^[12]	$Z(x_i) = Z^*(x_i) + \Delta(x_i)$	$Z^*(x_i)$ 为趋势项,辅助变量与目标变量之间的回归模型, $\Delta(x_i)$ 为残差项,使用克里金法进行预测。	建立降雨量与纬度、高程等的多元线性回归方程,对残差进行 OK 预测。

2 结果与分析

2.1 年尺度插值结果

对于 IDW 插值,影响它的因素主要是搜索策略的选择。表 3 中对比了多组幂指数 P 和邻近站点个数 n 的参数组合以便获得最佳的一组。由于该流域降雨空间差异明显,距离越近的对插值点的贡献更为显著和准确,因此 n 的取值不宜过大;然后通过固定 P 值,分别取 $P = 1$ 和 $P = 2$,改变 n 值比较插值结果;策略 IDW1 表示搜索区域形状为圆形,策略 IDW2 表示搜索区域形状为椭圆形,区别在于参数 Anisotropy factor 取值不同,分别为 1 和 0.2。结果表明:无论是从 MAE、RMSE 还是 CC 值,都可以看出 IDW2 组的插值结果明显好于 IDW1 组,最优的 IDW2($P = 2, n = 5$)比最优的 IDW1($P = 1, n = 7$) MAE 和 RMSE 分别低 11.63 和 19.08,即低了 16.7% 和 19.1%;而 CC 值提高了 8.2%。这是因为它的搜索区域形状为椭圆形,纬度跨度变小,这也间接说明年降雨量与纬度有良好的相关性(相关系数为 0.55),考虑空间异质性会明显提高插值结果精度。

另外,不同策略和 P 值下得到的插值最优结果的 n 值是不固定的, n 不是越大或者越小插值精度越高。除了 IDW2($P = 1, n = 7$)的 MBE 为正值外,其余的均小于零,说明 IDW 法对结果的预测值整体要小于观测值。

表 3 不同插值方法交叉验证结果

					mm
插值方法		MBE	MAE	RMSE	CC
IDW1	$n = 3$	-6.05	73.56	103.1	0.777
	$n = 5$	-3.69	69.80	100.6	0.788
	$n = 7$	-6.34	69.49	99.93	0.796
	$n = 9$	-7.35	73.01	104.0	0.780
	$n = 3$	-7.51	75.99	107.7	0.756
	$n = 5$	-5.73	71.46	105.1	0.768
	$n = 7$	-7.38	71.21	104.3	0.772
	$n = 9$	-7.08	73.06	106.3	0.764
IDW2	$n = 3$	-6.66	60.52	81.65	0.860
	$n = 5$	-3.17	59.21	82.38	0.858
	$n = 7$	2.03	61.63	85.56	0.848
	$n = 9$	-0.44	64.67	88.18	0.839
	$n = 3$	-7.44	60.70	81.28	0.860
	$n = 5$	-4.17	57.86	80.85	0.861
	$n = 7$	-1.01	59.03	82.29	0.857
	$n = 9$	-1.78	61.25	83.94	0.852
RBF	CRS	-11.17	62.84	88.33	0.840
	ST	-10.53	62.27	87.60	0.842
	IMQ	-6.68	60.16	84.77	0.850
OK	SPH	-3.09	59.20	84.85	0.848
	各向异性 EXP	-5.15	59.67	86.47	0.842
	GS	-3.19	62.21	88.44	0.835
	SPH	-2.13	63.46	91.49	0.825
	各向同性 EXP	-6.60	62.47	91.66	0.824
	GS	3.74	65.26	92.24	0.824
RK	各向异性 SPH	-0.17	54.13	76.07	0.877

对于 RBF 插值方法,参数取值根据 IDW 方法的经验,分别计算比较了 *CRS*、*ST* 和 *IMQ* 3 种径向基函数,结果见表 3。总体上讲,3 种函数的插值结果都比较理想,*CC* 值均大于 0.84,插值优劣顺序为 *IMQ* > *ST* > *CRS*。对于 *OK* 插值方法,在考虑方向性的基础上对比了球面、指数和高斯 3 种理论半方差模型。从 *MAE*、*RMSE* 和 *CC* 值可看出,对于同种模型,考虑各向异性好于不考虑各向异性的结果,3 种模型的插值优劣顺序为 *SPH* > *EXP* > *GS*,插值函数或模型的不同对 RBF 和 *OK* 法的精度影响不大。

RK 插值方法的多元线性回归在 *SPSS* 软件下进行。强制回归方法的结果表明参数间存在共线性^[12],故采用逐步回归的方法,剔除了经度、坡向和坡度 3 个变量,回归方程只包含纬度和高程两个变量,然后根据 *RK* 方法原理得到回归克里金模型,交叉验证结果见表 3。分别从 *IDW*、*RBF*、*OK* 和 *RK* 4 种插值方法选出最好的一组参数进行对比。从图 2 可以看出,不论 *MBE* 的绝对值、*MAE* 还是 *RMSE*,插值精度最高的都是 *RK* 方法,最低的是 *IDW1* ($P = 2, n = 7$),优劣排序为:*RK* > *IDW225* > *OK - SPH* > *RBF - IMQ* > *IDW213* > *IDW117* > *IDW127*;所有插值方法

MBE 均小于零,插值结果都低估了观测值。

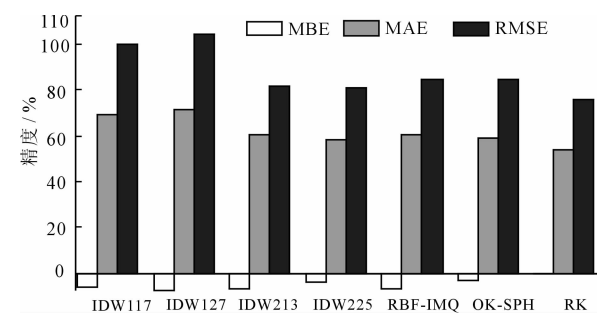


图 2 不同插值方法精度的比较

考虑了经度、纬度、高程、坡向和坡度的 *RK* 法精度最高,这是由于新安江流域属于山区,地理因子对降雨的影响较大。从回归方程也可以看出,年降水量随纬度的增加而减少,递减率为 821.8 mm/1°N,随着高程的增加略有增加,递增率 55.4 mm/100 m。

2.2 月尺度的插值结果

为了探讨不同插值技术对多雨期和少雨期的插值效果,基于年均插值方法的参数筛选结果,分别选取 4 种插值方法中精度最高的参数组合,使用 *MAE*、*RMSE* 和 *CC* 3 个指标对 48 个站点月均值雨量进行插值验证,结果见表 4。

表 4 基于月均值统计的插值方法比较 mm

指标	方法	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
MAE	IDW	4.52	5.28	5.62	5.96	8.44	10.44	10.57	12.10	6.51	4.37	3.00	3.72
	RBF	4.32	5.15	5.71	6.57	9.38	11.14	11.58	10.52	6.11	4.16	3.06	3.68
	OK	4.20	4.95	5.84	6.55	8.34	11.19	11.97	12.25	5.44	4.27	3.13	3.53
	RK	3.76	4.97	5.06	6.31	9.52	11.23	11.79	11.46	6.10	4.53	2.74	3.08
RMSE	IDW	6.98	7.61	7.46	8.03	12.08	15.70	14.43	15.39	8.26	7.04	4.18	4.77
	RBF	6.24	7.25	7.62	8.95	13.00	16.19	15.75	12.94	8.17	6.79	4.10	4.69
	OK	6.17	7.32	7.85	9.82	12.31	16.41	15.88	15.60	7.38	7.20	4.26	4.85
	RK	5.67	6.70	7.09	8.05	12.03	16.58	14.80	13.73	8.23	7.49	3.72	4.21
CC	IDW	0.614	0.882	0.811	0.930	0.922	0.797	0.773	0.624	0.557	0.541	0.791	0.647
	RBF	0.696	0.892	0.803	0.918	0.913	0.784	0.736	0.745	0.555	0.586	0.798	0.657
	OK	0.702	0.890	0.791	0.904	0.919	0.779	0.728	0.607	0.649	0.522	0.782	0.628
	RK	0.754	0.906	0.829	0.930	0.921	0.775	0.764	0.715	0.540	0.516	0.833	0.731

结合表 1 和表 4 可以得出,对于降雨量较大的月份(4 - 7 月,月均降雨量 > 180 mm),*IDW* 方法插值效果明显好于 *RBF* 和 *OK* 方法,略优于 *RK* 方法,说明该方法能捕捉住降雨极值点,而 *RBF* 和 *OK* 由于平滑效应显著使得精度下降^[2];对于降雨量较少的月份,*RBF* 和 *OK* 方法插值效果较 *IDW* 好,区别在于 *RBF* 对于降雨较均匀的月份插值效果好(8 月,10 月 - 12 月),而 *OK* 法对于空间自相关性高的数据,即符合正态分布的数据(1 月和 9 月)插值精

度高^[11,13]。*RK* 法的精度依赖于降雨量与地形因子的相关性:例如 4 月份,纬度与降雨量相关性为 0.77,经度与降雨量相关性为 0.53,插值拟合 *CC* 值高达 0.930;而 10 月份纬度与降雨量相关性仅为 0.21,经度为 0.06,插值 *CC* 值仅为 0.516;有研究表明^[7]只有当其与经纬度和海拔高程的相关系数大于 0.8 时考虑地理因素的方法才会优于 *IDW*,故在应用此方法前应该先做相关性分析。朱蕾^[12]等人对千岛湖周围降雨站插值结果认为,考虑经度、纬度

等地理地形因素的插值方法对非汛期的插值优于 OK 法。

从表 3 的月尺度分析结果还可以看出,不论采取何种插值技术,某些月份的插值精度都比较高,而某些月份较低。例如对于指标 CC 值,4 月份为 0.904~0.930,而 10 月份仅为 0.516~0.586。这说明插值精度不仅与插值技术选择有关,而且很大程度上取决于样本自身的空间自相关性等等。因此为了提高此类样本的插值精度,寻求和发展更为先进、合理的插值手段十分必要。

3 结 语

采用反距离权重(IDW)、径向基函数(RBF)、普通克里金(OK)和残差克里金(RK)4 种传统插值技术,对新安江流域 48 个雨量站的 11 年降雨数据进行插值研究,主要结论如下:

新安江流域降雨量在时间和空间上都存在较大的变异性,时间上存在丰雨季和少雨季,空间上降雨量与纬度相关性显著,因此考虑高程、纬度等地形因子影响的 RK 方法插值精度较高。参数筛选方面,对于 IDW 方法,搜索区域形状对插值精度影响较大,本文以椭圆形为佳,幂指数 P 和邻近站点个数 n 的取值对结果也有一定影响,应用时也要考虑;对于 OK 法,考虑各向异性的结果优于不考虑各向异性的结果,几种模型的插值效果都比较理想,应用时可以只选择常用的一种;插值方法具有平滑作用,且低估了观测值。月尺度的插值结果还表明,降雨量插值精度不仅与插值技术选择及参数设置有关,而且很大程度上取决于样本本身。

新安江流域作为皖浙两省的水源地保护区,对其研究和保护十分重要。同时由于该地区的降雨规律以及地形特点,对其他流域的插值研究具有一定的指导意义。首先,插值前应尽可能多的收集资料,特别是对于山区,要充分考虑高程、经纬度、坡向及坡度的影响;然后研究数据资料,确定样本的时空分布特征以及样本的自相关性,从而确定插值方法,最后根据所选择的插值方法筛选内部参数,获得最高的插值精度。本文以新安江流域为例,对其降雨量进行空间插值对比和参数筛选,以期为该流域水文

水质模型的应用提供可靠的降雨数据。除此之外,在对经典插值方法研究和改进的同时,非传统插值如信息熵和高精度建模插值技术的研究应用可能是未来的发展方向。

参考文献:

- [1] 张继国,谢 平,龚艳冰,等. 降雨信息空间插值研究评述与展望[J]. 水资源与水工程学报, 2012, 23(1): 6-9, 13.
- [2] Xie Yunfeng, Chen Tongbin, Lei Mei, et al. Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: Accuracy and uncertainty analysis [J]. Chemosphere, 2011, 82: 468-476.
- [3] Jin Li, Andrew D Heap. A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors [J]. Ecological Informatics, 2011, 6: 228-241.
- [4] Piazza A Di, Conti F Lo, Noto L V, et al. Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy [J]. International Journal of Applied Earth Observation and Geoinformation, 2011, 13:396-408.
- [5] 曾红伟,李丽娟,张永莹,等. 大样本降水空间插值研究:以 2009 年中国年降水为例[J]. 地理科学进展, 2011, 30(7): 811-818.
- [6] 季 青,余 明. 基于协同克里格插值法的年均温空间插值的参数选择研究[J]. 首都师范大学学报(自然科学版), 2010, 31(4): 81-87.
- [7] 封志明,杨艳昭,丁晓强,等. 气象要素空间插值方法优化[J]. 地理研究, 2004, 23(3): 357-364.
- [8] George Y Lu, David W Wong. An adaptive inverse-distance weighting spatial interpolation technique [J]. Computers & Geosciences, 2008(34):1044-1055.
- [9] 杨雨亭,尚松浩,李 超. 土壤水分空间插值的克里金平滑效应修正方法[J]. 水科学进展, 2010, 21(2): 208-213.
- [10] 徐 英. 考虑块段效应的 RBF 神经网络在土壤空间插值中的应用[J]. 水科学进展, 2012, 23(1): 67-73.
- [11] 汪翠英,杨勤科,郭伟玲,等. 区域水土流失中降雨量空间插值方法选取[J]. 水土保持研究, 2008, 15(2): 88-91.
- [12] 朱 蕾,黄敬峰. 山区县域尺度降水量空间插值方法比较[J]. 农业工程学报, 2007, 23(7): 80-85.
- [13] 刘登伟,封志明,杨艳昭. 海河流域降水空间插值方法的选取[J]. 地球信息科学, 2006, 84:75-79+83.