

降雨资料 Kriging 与 IDW 插值对比分析 —以漓江流域为例

范玉洁^{1,2}, 余新晓¹, 张红霞², 宋美华²

(1.北京林业大学水土保持学院, 北京 100083; 2.桂林理工大学环境科学与工程学院, 广西 桂林 541004)

摘要:降水空间化信息在很多领域都具有重要意义,而进行空间插值方法形成降水空间化信息是当代较为常用的方式。面对众多的插值方法其插值精度成为我们是否采用的关键,就 kriging 插值法与 IDW(Inverse Distance Weighting)插值进行研究以探讨其插值效果,为此类科学研究提供依据。研究以漓江流域内各气象站点降水统计资料为基础分别采用上述两种方法进行插值处理,通过与预留实测站点比较评判方法的插值效果的优劣。结果表明,多年月平均降水量作为时间步长时,降水丰沛的月份使用 kriging 插值法较优于 IDW 插值法,而枯水月份则使用 IDW 插值法较优于 kriging 插值法。

关键词:kriging 插值;IDW 插值;降水量

中图分类号:TV125

文献标识码:A

文章编号:1000-0852(2014)06-0061-06

1 引言

降水对人类社会的发展有着非常重大的影响,其空间化信息在很多领域都具有重要意义。例如生态环境治理、区域水文、水资源分析及管理、旱涝灾害管理。通常情况下,我们获取降雨数据的方式都是通过自动雨量观测站或者分散的气象台站,然而受各种因素的限制,布设的水文站点往往是有限的,并且有限的站点在空间上的布局也不全是合理的。对于地形复杂的偏远地区,获取多年气象资料难度较大;因此,目前获取某流域降雨量信息的一个行之有效的办法就是根据区域内部分站点的实测降雨数据进行空间插值^[1]。

空间插值是基于一部分所获取的能体现该区域某些地理要素的空间分布全部特征或部分特征的空间样本,用这些已知的地理要素未知的空间特征,即利用已知的空间数据预测未知空间得到预测数据值;其本质是运用少部分已知点对完整的未知区域进行空间预估^[2]。空间插值的方法多种多样,目前国内外常见的插值方法有反距离加权法 (Inverse Distance Weight tension

IDW)、张力样条函数法(Spline With Tension)、趋势面法(Trend)、普通克里金法(Ordinary Kriging)、协同克里金法(Ordinary Cokriging)、全局多项式插值(Global Polynomial Interpolation)、局部多项式插值 (Local Polynomial Interpolation)、径向基函数(Radial Basis Functions)等^[3-4]。

许多学者比较分析了关于不同区域的降水数据的空间插值方法。例如,Tabios 和 Salas 对时间跨度较长的月累计降水量进行了研究,结果表明基于地统计学的 kriging 法效果比常用的距离权重倒数法和泰森多边形法好^[5];Hogg 在对日累计降水量进行插值后,其也得到了相似的结果^[6];Dirks 使用四种不同的方法对多种时间步长的高分辨率降水数据进行空间插值,结果显示:综合比较计算的需要用时及方法的繁琐程度,距离权重倒数法的插值精度优于其他方法;Parajka 采用 Kriging 插值法和 Ordinary Cokriging 法(此法考虑了地形因素)对不同地区进行了降水数据空间插值,结论是:当站点密度较大时,kriging 法的插值效果更让人满意,然而在一些偏远的山区,由于站点较少,此时考虑了地形因素的 Ordinary Cokriging 法效果更好^[7]。彭波,钟仕全,苏旭芳等^[8]在对漓江流域降雨量空间插值

收稿日期:2013-01-28

基金项目:国家“十二五”科技支撑项目(2011BAD38B05);林业公益性行业科研专项(201104005)

作者简介:范玉洁(1974-),男,广西桂林人,博士,主要从事水土保持方面研究。E-mail: fanyujie@glite.edu.cn

通讯作者:余新晓(1961-),男,甘肃平凉人,教授,博士生导师,主要从事生态水文、水土保持方面的研究。E-mail: yuxinxiao111@126.com

精度分析中,对漓江流域周边 15 个气象台站 2010 年 6 月夏季全年 3 个时间段降雨量数据进行统计分析分别运用了 IDW 法、Kriging 插值法、Ordinary Cokriging 法进行插值处理并用流域内 15 个自动站对应的降雨量数据进行精度验证。何艳虎,林凯荣^[9]在降雨空间插值方法在东江流域的比较运用中,使用 ArcMAP 提供的三种空间插值方法(IDW 法、Spline 法和 Kriging 法),以东江流域 46 个站点的已知的实测降雨量为数据源,且在雨量站的数目选用方面做了改变,以此进行降雨量的空间插值并验证了各项插值成果;结论为:对东江流域降雨量的空间插值,相比较而言,IDW 法更优于其他。而其他三种插值方法的插值结果在一定程度上也基本体现了区域降雨量的空间分布特性;且研究也表明流域内已知站点的数目对东江流域的降雨数据空间插值存在影响;不同插值方法的选择会得到不同的插值效果。瑞士洛桑大学地球科学研究所的空间插值 97 比较计划,利用瑞士的降水观测资料系统地比较和分析了多种插值方法之间的差异性和不确定性。还有许多研究人员也使用气象观测数据在不同的地域、不同的时间步长内进行了各种数值分析,包括空气温度、土壤温度及湿度、降水等,运用多种现有的插值方法做了多种相关的比较和研究。受各种条件的限制,如高质量的基础数据和专业基础,致使这些类研究的可操作性与实用性都略显不足。而绝大部分的研究所得出的结论都是:空间数据的插值,不存在哪一种最优的空间内插法;一种适用于任何量值的空间插值也是不存在的;对于不同的空间变量,在不同的地域和不同的时空尺度内所谓的最优内插法都是相对的^[10]。

而事实上不可避免的是:对于相同的数据,使用不同的插值方法所得到的结果可能会存在较大的差异。这是因为所研究区域的地形地貌等特征不同,以及研究的数据不同,并且降水数据时空分布上的不均匀、趋势性、周期性等特征导致的。因此,对降水量这一方面而言,众多科研人员试图采用各种方法研究降水序列,并使用 ArcGIS 软件中的 AcrMap 对区域内或外离散的数据,对空间进行插值,以期得到降水量在整个研究区域的降水空间分布结果。而这类研究,已成为当前国内外的科学研究热点。

随着水文模型和水利计算对降雨精度要求的不断提高,有必要通过不同插值方法的比较,找出适合流域降雨量的空间插值方法。但对于空间插值对比分析这

方面,有关研究对气温插值的关注较多,对降水则较少。降水量的空间插值不确定因素更多,所以有必要进行有针对性的分析,从而减少不确定因素。本文试图以桂林漓江流域为例,对 kriging 法与 IDW(Inverse Distance Weighting)插值方法进行分析以考量二者插值精度,为降水空间插值研究提供参考。

2 研究方法

本文以漓江流域周边县市 17 个气象台站 1961~2010 年共 50 年的逐日降水统计资料为基础,综合考虑所研究内容与漓江流域降水时空分布特性选取研究时段。因漓江流域 5~8 月属于区域降水相对比较集中且丰沛的时期,降水均匀、数值也相对较大,这可以更好地体现插值结果,插值效果也更加明显,便于进行分析对比。而 1 月属于最枯月份,对其进行插值,可用于与降水丰沛的 5~8 月插值结果进行比较。故从降水统计资料中分别汇总出各站点 50 年平均 1 月、5 月、6 月、7 月、8 月的月降雨量。运用基于 ArcGIS 的两种空间插值方法进行插值计算(距离反比插值法采用二阶方程,kriging 法采用球形模型进行运算)。选取位于漓江流域内的桂林及融安 2 个站点为检验点,用平均绝对误差(MAE)和插值平均误差平方的平方根(RMSE)两指标考量二者精度。

2.1 Kriging 插值法

Kriging 插值法又称空间自协方差最佳插值法,由南非地质学家克立格(D.G.Krige)于 20 世纪 50 年代提出,1962 年法国学者 G.Matheron 第一次引入区域化变量的概念,将 Kriging 插值法理论化、系统化,进行了进一步的推广和完善^[11]。

它的主要优点是通过对整组空间数据点的关系来进行空间插值,从而更加真实、也更加客观,同时预测区的误差范围大小也可由误差等值线确定。缺点是所需要大量的空间数据点,从而增加了调查的成本,另外变异函数需要根据经验人为选定。

计算公式为:

$$Z(\chi_0) = \sum_{i=1}^n \lambda_i Z(\chi_i) \quad (1)$$

式中: $Z(\chi_0)$ 为 χ_0 处降雨量估计值; $Z(\chi_i)$ 为 χ_i 处的观测值; λ 为 kriging(克里格)权重系数; λ_i 为 χ_i 处的 kriging(克里格)权重系数; n 为观测点个数。

Kriging 插值算法中提供了许多种半变异函数模型,如高斯、线形、球形、阻尼正弦和指数模型等,而相

关研究表明在对气象要素场插值时使用球形模型效果比较好。

2.2 IDW(Inverse Distance Weighting)插值法

IDW 插值算法,也称为反距离权重法,是最常用的空间插值方法之一,该方法由美国国家气象局于 1972 年提出。该法用周边相邻的采样点的值,估计未知点的值,以待插点与实际观测样本点之间的距离为权重,离插值点越近的样本点赋予的权重越大,其权重贡献与距离成反比。该方法是基于相近相似的原理:即两个物体离得远,它们的性质相似性就越小,反之,物体间离得越近则相似性越大。

计算公式为:

$$Z(\chi_0) = \frac{\sum_{i=1}^n Z(\chi_i) \cdot w_i}{\sum_{i=1}^n w_i} \quad (2)$$

式中: $Z(\chi_0)$ 为待估计的 χ_0 点属性值; $Z(\chi_i)$ 为 (χ_0) 点周围局部领域内第*i*点 χ_i 的属性值; n 为局部邻域内点的个数; w_i 为 χ_i 点对于 χ_0 点的权值。

反距离权重法主要依赖于反距离的幂值。幂参数可以通过输出点的距离来控制已知点对空间内插值的影响。幂指数越大,较远的点对于输入的影响越小,即幂指数越高,其局部影响的程度越高。幂参数是一个正实数,可根据均方根误差最小的选择标准,采用逐步迭代的方法对研究区域插值量的幂指数进行筛选,进而确定出适合的使用值,默认值为 2。

2.3 检验方法

由于评价空间插值精度比较困难,目前通常采用交叉验证法来进行验证。通过计算样本实测值与估算值两者间的误差来评估插值方法的精度。本文采用常用的平均绝对误差(MAE)和平均误差平方的平方根(RMSE)来反映 IDW 插值法和 kriging 插值法的空间插值能力。平均绝对误差可以评估估算值可能的误差范围。平均误差平方的平方根可以反映利用样点的估算灵敏度和极值效应。MAE 和 RMSE 越小说明插值结果越好,反之,则越差。

计算公式为:

$$MAE = \frac{1}{n} \sum_{i=1}^n (|O_i - E_i|) \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - E_i)^2}{n}} \quad (4)$$

式中: O_i 表示第*i*站的实测降水量; E_i 表示第*i*站的预测降水量; n 表示用于检验的站点数。

3 数据处理

3.1 研究区概况

漓江流域地处北纬 24°18'~25°41',东经 109°45'~110°40',位于广西壮族自治区的东北部、南岭山地西部、湘桂走廊西南端,平均宽度约 34km,南北长度约全长 214km,流域面积为 6 050km²。漓江流域受太平洋副热带高压西侧的偏南气流影响,其上空时常水汽丰沛;属于山区丘陵雨源性河流,虽径流年内分配极不均匀,但雨量充沛,水量丰富。漓江流域年总降水量为 108×10⁸m³,其中蒸发量为 28×10⁸m³,流出水量为 70×10⁸m³,地下水补给量为 10×10⁸m³[12]。漓江 3~8 月为丰水期,9 月~次年 2 月为枯水期,其中 5~6 月占全年径流量的 40%,枯水期占全年径流量的 20%,而最枯时期 1 月仅占全年径流量的 2%。

此次插值数据来源于漓江流域周边县市 17 个气象台站(见表 1)的 1961~2010 年共 50 年的逐日降水统计资料。由统计资料计算出所选时段各站点多年平均月降水量(见表 2)。

表 1 站名及位置

Table 1 The station names and positions

站名	东经 / °	北纬 / °
风山	107.033	24.550
河池	108.050	24.700
都安	108.100	23.933
柳州	109.400	24.350
蒙山	110.517	24.200
贺县	111.517	24.417
平果	107.583	23.317
来宾	109.233	23.750
桂平	110.083	23.400
梧州	111.300	23.483
南宁	108.350	22.817
灵山	109.300	22.417
玉林	110.167	22.650
钦州	108.617	21.950
北海	109.117	21.483
融安	109.400	25.217
桂林	110.300	25.317

3.2 插值数据分析

因 kriging 方法是建立在平稳假设的基础上,这种假设在一定程度上要求所有数据值具有相同的变异

表2 多年平均月降水量(mm)
Table 2 The mean monthly precipitation (mm)

站名	1月	5月	6月	7月	8月
风山	24.418	210.912	317.716	298.642	231.722
河池	34.554	217.364	293.980	245.018	220.094
都安	41.912	249.716	361.270	306.972	247.134
柳州	47.286	233.906	260.268	191.584	175.732
蒙山	60.984	293.392	329.232	227.878	197.216
贺县	67.644	264.858	256.312	165.676	164.902
平果	41.113	179.400	104.062	255.717	185.291
来宾	44.062	213.108	554.452	212.170	175.352
桂平	63.428	280.702	285.216	228.968	193.716
梧州	54.012	249.600	246.572	177.258	180.996
南宁	37.644	179.092	264.274	517.616	198.412
灵山	47.206	197.662	485.928	288.990	272.228
玉林	53.912	236.362	183.574	223.932	225.626
钦州	45.366	216.790	404.980	444.384	400.484
北海	32.974	140.778	290.318	364.414	392.520
融安	57.428	312.952	421.856	265.688	179.216
桂林	60.722	334.138	366.766	227.196	154.570

数据的分位数 10^{-1}

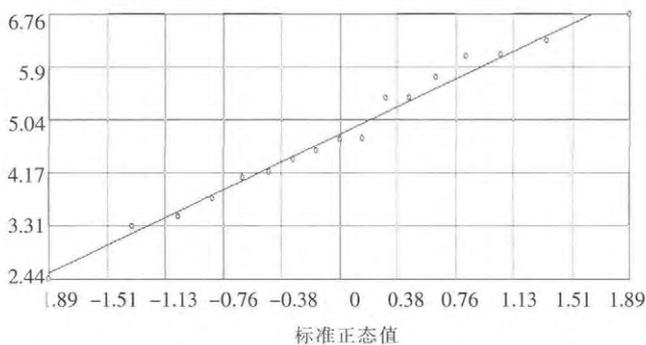


图1 多年平均1月降雨量QQPlot分布图

Fig.1 QQPlot distribution of the mean annual rainfall in January

数据的分位数

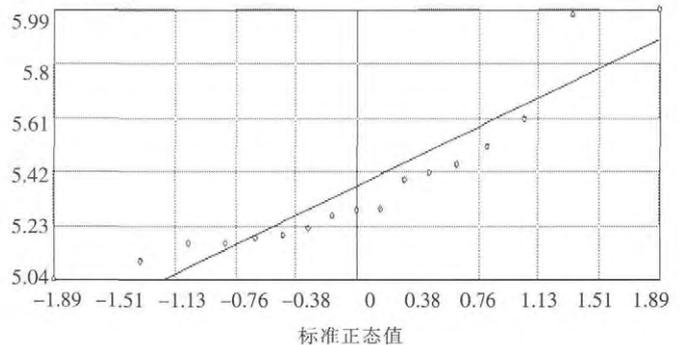


图2 多年平均8月降雨量QQPlot分布图

Fig.2 QQPlot distribution of the mean annual rainfall in August

性。另外,一些 kriging 插值(如 ordinary kriging 法、simple kriging 法和 universal kriging 法等)都假设数据服从正态分布。如果数据不服从正态分布,则可进行一定的数据变换,从而使其服从正态分布。因此,在进行统计分析前,需要对数据进行检查是否符合正态分布,检验数据分布特征。本文通过 ArcGIS 地统计分析模块中所含的探索性空间数据分析工具 Normal QQPlot(正态 QQPlot 分布图)完成。得正态 QQPlot(Normal QQ-Plot)分布图,其中 5、6、7 月图形与 1 月份(图 1)类似,8 月份区别较大,但所有研究时期内,从图形可知所选

插值时段数据都近似为一条直线,所以可以进行 kriging 插值计算。对于 8 月份图 2 中尾部的两个离散点,可能是地形、地貌或者地区性气候等的差异造成真实异常值,也可能是不正确的测量或记录统计导致^[5]。一般而言,在插值分析计算前,如果存在离群值来自于测量或者数据输入的明显错误,那么就要改正或删除它们。通过对所选用数据进行仔细检查,插值分析中所用的降水数据均来自区域气象站的观测数据。即便这些数据中可能存在误差,这些误差可能来源于仪器误差或者人为因素,但由于缺乏其它相关来源的数据对其

表 3 不同插值方法多年平均月降水量预测值与误差

单位:mm

Table 3 The forecasted mean monthly rainfall and errors using different interpolation methods (mm)

站点	月份	实测值	预测值		误差	
			kriging 插值法	IDW 插值法	kriging 插值法	IDW 插值法
融安	1	57.43	49.49	52.37	-7.94	-5.05
	5	312.95	275.78	271.08	-37.17	-41.87
	6	421.86	313.69	301.00	-108.17	-120.86
桂林	7	265.16	220.60	213.65	-45.09	-52.03
	8	179.22	176.37	169.93	-2.84	-9.29
	1	60.72	61.68	57.27	0.96	-3.45
林	5	334.13	301.79	287.35	-32.34	-46.79
	6	366.77	378.47	304.26	11.71	-62.50
	7	227.20	235.38	238.48	8.18	11.28
	8	154.57	183.41	182.20	28.84	27.64

表 4 多年平均月降水量插值分析 单位:mm

Table 4 The interpolation analysis of the mean monthly precipitation (mm)

方法		月份	Kriging 插值方法	IDW 插值方法
检验指标				
MAE		1	4.49	1.98
		5	34.76	44.33
		6	48.23	91.68
		7	18.45	20.37
		8	13.00	9.17
RMSE		1	7.05	2.46
		5	34.84	44.39
		6	76.93	96.21
		7	32.41	37.65
	8	20.49	20.61	

进行校验和订正。且考虑到各站点的数据都是获取于标准气象台站的观测资料,因而其可靠性应该得到保证,在插值处理时不特别对离群点和引起趋势变化大的点进行剔除或修改。

在上述分析基础上获得预留检验站点(桂林和融安),插值结果(表3)与相应误差。由插值计算值求得 Kriging 与 IDW 插值平均绝对误差(MAE)和插值平均误差平方的平方根(RMSE)两个检验指标值(表4)。

由表4可知,丰水期除8月 kriging 插值法对于各月均值插值结果的 MAE 值与 IDW 插值法所得的值相近外,其余5、6、7月 kriging 插值的 MAE 值均小于

IDW 插值。而所有插值的丰水期月份使用 kriging 插值法获得的 RMSE 值均小于 IDW 插值法的值。综合比较可知,对于降水丰沛的月份使用 kriging 插值法较优于 IDW 插值法。而最枯水月1月 kriging 插值的 MAE 值和 RMSE 值均大于 IDW 插值。说明枯水月份使用 IDW 插值法较优于 kriging 插值法。同时从表中也可得出对于不同时期的降水量空间插值,较优的插值方法也是不同的。

4 结论与讨论

本文采用漓江流域县市17个站点1961~2010年的降水观测资料,对 kriging 插值法和常用的 IDW(Inverse Distance Weighting)插值法两种空间插值方法进行考量得如下结论:

对于使用多年月平均降水量作为时间步长时,kriging 插值与 IDW 插值方法的最优选择与降水量的丰沛程度多少有关。降水丰沛的月份使用 kriging 插值法较优于 IDW 插值法,而枯水月份则使用 IDW 插值法较优于 kriging 插值法。

因时间等客观原因的限制,本文未进一步分析造成这种相反结论的原因。由于此次分析研究并未选取其他的时间步长、更多的站点以及这两种方法之外的其他方法对数据进行空间插值,而插值过程中站点数量、气象站点分布状况、观测资料本身误差、插值方法的选择、经纬度、高程、坡度坡向、海陆距离、下垫面状况、大气环流差等都会对插值精度存在或多或少的影 响。所以,要想获得更精确的插值效果,还必须引入更多的分析因子。

总而言之,空间插值方法各种方法都有其优缺点,其适用范围、算法用法也都有差异。可以肯定的是,对于特定的空间特定的时空尺度,没有最优的插值方法,只有相对而言较优的,选用插值方法时,应因地制宜。

参考文献:

- [1] 邹强.流域降雨空间插值方法比较[J].节水灌溉,2009,(7):12-14.(ZOU Qiang. Comparison of river basin rainfall's spatial interpolation method [J]. Water Saving Irrigation, 2009,(7):12-14.(in Chinese))
- [2] 董晓华,薄会娟,邓霞,等.降雨空间插值方法及在清江流域的应用[J].三峡大学学报:自然科学版,2009,31(6):6-10.(DONG Xiaohua, BO Huijuan, DENG Xia, et al. Rainfall spatial interpolation methods and their applications to Qingjiang River basin[J]. Journal of China Three Gorges University (Natural Sciences), 2009,31(6):6-10. (in Chinese))
- [3] 汤国安,杨昕.ArcGIS 地理信息系统空间分析实验教程[M].北京:科学出版社,2006.(TANG Guo-an, YANG Xin. ArcGIS Experiment Course of GIS Geographic Information System's Spatial Analysis[M]. Beijing: Science Press, 2006. (in Chinese))
- [4] 罗明,裴韬.空间软数据及其插值方法研究进展[J].地理科学进展,2009,25(5):663-672.(LUO Ming, PEI Tao. Review on soft spatial data and its spatial interpolation methods[J]. Progress in Geography, 2009,25(5):663-672. (in Chinese))
- [5] Tabios III, G Q, Salas J D. A comparative analysis of techniques for spatial interpolation of precipitation [J]. Water Resources Bulletin, 1985,21(3):265-380.
- [6] Bussires, W Hogg. The objective analysis of daily rainfall by distance weighting schemes on a esosca legrid[J]. Atmosphere-Ocean, 1989,27(3):521-541.
- [7] Parajka J. Mapping long-term mean annual precipitation in Slovakia using geostatistical procedures [A]. Vlasak P, Filip P, Chara Z. Proceedings of the International Conference on Problems in Fluid Mechanics and Hydrology [C]. Prague: Institute of Hydrodynamics, Academy of Sciences, 1999:424-430.
- [8] 彭波,钟仕全,苏旭芳,等.漓江流域降雨量空间插值精度分析[J].气象与研究与应用,2011,32(4):30-33.(PENG Bo, ZHONG Shiquan, SU Xufang, et al. Analysis on rainfall spatial interpolation precision in Lijiang River basin [J]. Journal of Meteorological Research and Application, 2011,32(4):30-33. (in Chinese))
- [9] 何艳虎,林凯荣.降雨空间插值方法在东江流域的比较运用[J].水力发电,2010,36(10):7-9.(HE Yanhu, LIN Kairong. Application and comparison of rainfall spatial interpolation methods on Dongjiang River basin [J]. Water Power, 2010,36(10):7-9.(in Chinese))
- [10] Haining R. Spatial Data Analysis in the Social and Environmental Sciences [M]. Cambridge: Cambridge University Press, 1990:291-312.
- [11] 彭楠峰.距离反比插值算法与 Kriging 插值算法的比较[J].大众科技,2008,105(5):57-58.(PENG Nanfeng. Comparison of distance-inverse interpolation method and Kriging interpolation method [J]. Popular Science & Technology, 2008,105(5):57-58. (in Chinese))
- [12] 成官文.漓江流域水环境综合治理对策[J].南方国土资源,1996,(6):27-30.(CHENG Guanwen. Synthesis prevent and bring Lijiang River valley under control[J]. Southern Land and Resources, 1996,(6):27-30.(in Chinese))

Comparison between Kirging Interpolation Method and Inverse Distance Weighting Tension for Precipitation Data Analysis: Taking Lijiang River Basin As A Study Case

FAN Yujie^{1,2}, YU Xinxiao¹, ZHANG Hongxia², SONG Meihua²

(1. College of Soil and Water Conservation, Beijing Forestry University, Beijing 100083, China; 2. School of Environmental Science and Engineering, Guilin University of Technology, Guilin 541004, China)

Abstract: Precipitation spatial information is very important in many fields, for which the various spatial interpolation methods are usually used. The paper compared Kirging interpolation method and IDW (Inverse Distance Weighting) to analyze the interpolation results. The two method were used to make interpolation processing the precipitation data from the meteorological stations in the Lijiang River Basin, and to discuss their relative merits and applicable conditions by comparing the observed and interpolated values. The results show that when the mean monthly precipitation is regarded as time step, Kriging interpolation method is better than IDW in months with abundant precipitation, whereas, IDW is prior to Kriging interpolation method in dry months.

Key words: Kirging interpolation method; inverse distance weighting tension; precipitation