

江西多年平均降水量空间插值模型的选取与比较

张余庆¹, 陈昌春¹, 尹义星², 杨绪红³

(1. 南京信息工程大学 遥感学院, 南京 210044;

2. 南京信息工程大学 水文气象学院, 南京 210044; 3. 南京大学 地理与海洋科学学院, 南京 210093)

摘要:为了探究江西省多年平均降水量的空间分布格局, 本文选取 81 个时间序列完整的气象站 30 a(1976—2005 年)降水数据, 运用泛克里格中不同半变异模型对降水数据进行拟合。分别采用 65 个建模站点交叉验证和 16 个验证站点的检验, 证实了该研究区域多年平均降水量存在较强的空间相关性。通过不同半变异函数模型的对比发现, 球面和指数模型在建模站点交叉验证结果中的标准均方根预测误差分别为 1.024 和 1.023, 较为接近 1, 表明其误差较小; 其在验证站点检验结果中的标准均方根预测误差分别为 1.105 和 1.104, 表明这两种模型的拟合效果较优, 能较为真实地反映江西省多年平均降水量的空间分布情况, 其中以指数模型拟合效果最优。

关键词:江西省; 降水量; 空间插值模型; 克里格插值

中图分类号: P333.1

文献标识码: A

文章编号: 1005-3409(2013)04-0069-06

Spatial Interpolation Model Selection of Multi-Year Average Precipitation in Jiangxi Province

ZHANG Yu-qing¹, CHEN Chang-chun¹, YIN Yi-xing², YANG Xu-hong³

(1. College of Remote Sensing, Nanjing University of Information Science & Technology, Nanjing

210044, China; 2. College of Hydrometeorology, Nanjing University of Information Science & Technology,

Nanjing 210044, China; 3. College of Geographic and Oceanographic Sciences, Nanjing University, Nanjing 210093, China)

Abstract: To explore the distribution pattern of multi-year average annual precipitation in Jiangxi Province, 81 complete sequence time series of 81 weather stations were used, which included precipitation data of 30 years (1976—2005) in Jiangxi Province. The universal kriging is used in different semivariogram model to fit the precipitation data after exploring the distribution features of the data; according to test of 65 cross validation and inspection of 16 stations, it is indicated that the multi-year average annual precipitation has strong spatial correlation. Through the comparison of different semivariogram function models, it is found that exponential model and spherical model in the results of cross validation stations were 1.024, 1.023 for Root-Mean-Square Standardized, which is close to 1 and error was small. Exponential model and spherical model in the results of inspection stations were 1.105, 1.104 for Root-Mean-Square Standardized, showing that the fitting effect of the two models is better than the others, it can more perfectly reflect the spatial distribution of multi-year average precipitation in Jiangxi Province, in which the result of exponential model can obtain the better integer effect within the existing weather stations network.

Key words: Jiangxi province; precipitation; spatial interpolation model; kriging interpolation

降水是我国水资源的重要来源, 对工农业生产和建设具有重大意义。由于人力、物力和财力有限, 气象站点布设不可能区域全覆盖, 致使大部分区域空间降水数据无法直接获取^[1], 故降水量空间插值研究对

水资源的合理评估具有重要意义, 也成为当前的研究热点和难点^[2-7]。由于降水量存在时间与空间上的变异性和不确定性, 故对其进行空间插值拟合研究较多, 如胡庆芳等^[8]对 2003—2009 年赣江流域降水空

间变异规律进行分析,采用全局半方差函数描述流域降水的整体相关结构,并用局部 Moran's I(莫兰)指数揭示了降水局部聚集形态和地带性规律;朱会义等^[9]基于 1990 年潮白河流域 58 个雨量站的降雨观测数据,采用不同插值方法对其进行降雨空间不确定性分析,研究表明,观测站点数量越多,区域降雨插值的不确定性越小,时间尺度越短降雨空间插值的不确定性越大;此外,由于影响降水空间分布的因素不同,所选择的降水插值方法和降水拟合模型也不尽相同^[10]。而气象站点的个数、空间分布以及插值方法对研究区域降水量插值精度影响较大,在站点数量和空间分布相同的条件下,空间插值方法成为影响区域降水空间插值精度的主要因素^[11]。江西省降水量的时空分布比较集中,赣西北地区的降水集中期平均每年只有赣东北的 50%,降水集中期的主要时段为 6 月中旬至 7 月上旬,此时段也是洪涝灾害频发的高峰期,给工农业生产造成极为严重的损失^[12]。且在已有的文献中,有关江西省多年平均降水量空间插值的深入研究并不多见。鉴于此,为了探求江西省多年平均降水量的空间分布状况,本文利用 ArcGIS 对 1976—2005 年江西省 30 a 间 81 个气象站点年均降水量进行泛克里格插值,采用建模站点交叉验证及验证站点来检验对比不同模型半变异函数的插值精度情况,以期为江西省现有观测站点分布背景下的降水空间分布特征研究和区域工农用水决策及洪涝灾害防护提供科学依据。

1 研究区概况与数据来源

江西省位于中国东南部,长江中下游南岸,地处北纬 $24^{\circ}29'$ — $30^{\circ}04'$ 、东经 $113^{\circ}34'$ — $118^{\circ}28'$,土地总面积 $1.67 \times 10^5 \text{ km}^2$,居华东地区六省一市之首,是长江三角洲、珠江三角洲及海西经济区等发达地区的共同腹地;该区域属亚热带湿润季风气候区,气候温暖,多年平均气温 18°C ,全省雨量充沛,多年平均降水量为 $1\ 635 \text{ mm}$,总体分布格局呈现为南多北少、东多西少、山区多盆地少,6—7 月降水量较多,是洪涝灾害频发的高峰期;江西常态地貌类型以山地、丘陵为主,分别占全省总面积的 36% 和 42%;全省水资源丰富,总量约为 $1\ 505 \text{ 亿 m}^3$,人均拥有量高于全国平均水平,大小河流 2 400 多条,总长约 1.84 万 km ,其中赣江、抚河、信江、修水及饶河为全省五大江河,位于江西北部的鄱阳湖是中国目前最大的淡水湖泊和世界最大的候鸟栖息地。

本研究涉及的降水和站点经纬度数据来源于江西省气象局。研究区域内共有 87 个气象站,但考虑

到数据的完整性、可靠性和代表性(部分站点缺测、观测年限较短以及分布非均匀)经整理后选取 81 个站点 1976—2005 年 30 a 间逐日降水观测数据作为本文的研究数据。选取 65 个站点数据进行建模,16 个站点数据用于验证,添加投影系统(本文采用 WGS_1984_Xian_1980_3_Degree_GK_CM_114E 投影坐标系),建立降水空间数据库。各个气象站点的空间分布如图 1 所示。

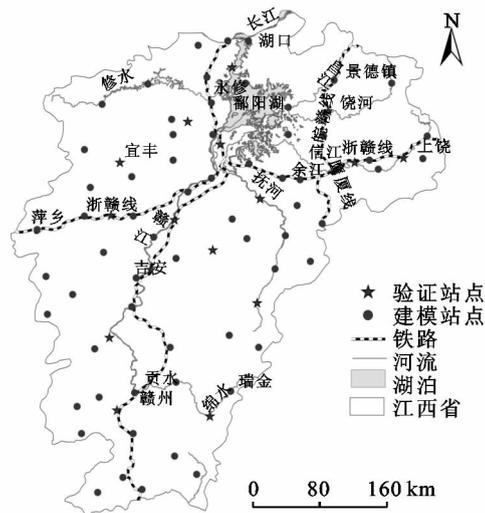


图 1 江西省气象站点分布

2 数据预处理

2.1 空间自相关性分析

空间自相关是按照空间赋值状况测量各变量值之间的相关关系,若相似的值在空间上相互靠近,则被描述为相关;反之,则为独立或随机分布。空间临近区域单元的相似程度可根据 Moran's I 来判断, Moran's I 取值范围为 $-1 \sim 1$,越接近 -1 代表单元间属性值差异越大,越接近 1 代表单元间属性值关系越密切,性质越相近,接近 0 代表单元间属性值不相关^[13]。本文基于 ArcGIS 相关工具采用反距离加权的欧几里得距离(Euclidean Distance)法,算得 30 a 年均降水量 Moran's I 值为 0.65 ,说明江西省降水数据具有较高的空间自相关性。

2.2 数据分布检验

进行降水数据插值之前,需依据数据统计分布特征选取合适的插值方法,常用的数据统计特征检验方法有直方图法、正态 QQ 分布图法和趋势图法等^[14]。一般的,数据统计分布特征常出现数据聚集以及离群值等现象,因此通常需采取 \log (对数)变换和 Box-Cox(博克斯—考克斯)变换进行数据预处理。

研究区各气象站降水量的基本统计表明:江西省总体年均降水量为 $1\ 430.7 \sim 1\ 973.1 \text{ mm}$,均值为

1 653.8 mm, 中位数与年降水量平均值非常接近, 为 1 648.7 mm; Skewness(偏态系数)为 0.213 3, 呈现正偏; Kurtosis(峰度系数)为 2.273 7, 表明该区域降水数据的分布比标准正态分布高耸且狭窄, 更集中于平均值附近(图 2a)。图 2b 正态 QQ 分布分析表明, 数据总体分布倾向于正态分布, 故无需进行数据变换。图 2c 空间趋势分析表明, 该区域降水量具有较

强的空间相关性(X 轴正向为东, Y 轴正向为北, Z 轴为降水量), 其基本趋势为东高西低, 且降水量与经度有较强的相关性, 南北向各地区降水差异较大, 呈现为倒“U”形变化趋势。该降水量的空间特征与江西省地势有关, 该省东、南、西三面环山, 东部山脉高程较低, 夏季东南季风带来丰富的水汽, 形成了较高的降水量。

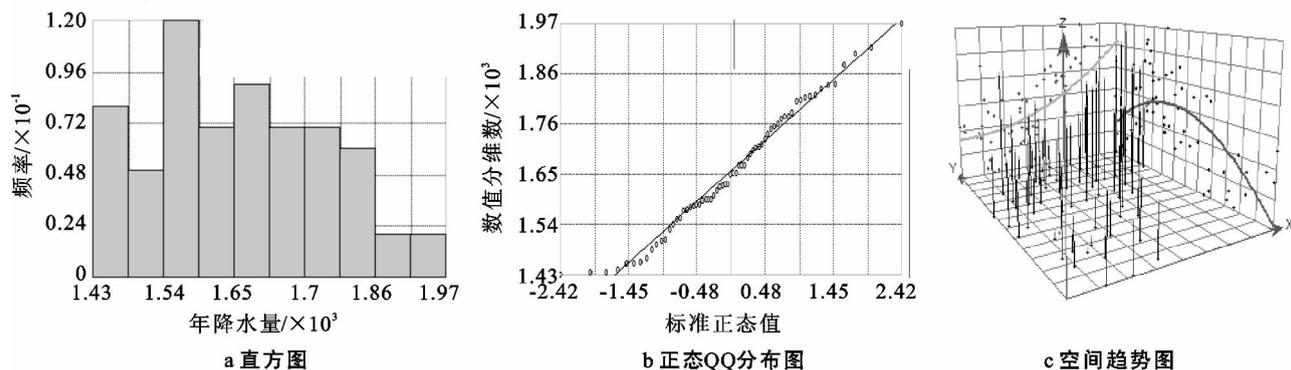


图 2 建模站点数据检验分析

3 降水空间结构关系

3.1 降水空间变异性

半变异云图是探测研究对象空间变异性的重要工具, 能够定量描述各站点之间的空间相关程度。降水量的半变异函数 $r(h)$ 选用公式(1):

$$r(h) = \frac{1}{2n} \sum_{i=1}^n [z(x_i) - z(x_i + h)]^2 \quad (1)$$

式中: $r(h)$ ——按步长 h 归类样点的平均半变异; n ——按方向归类的样点对个数; z ——区域变化量; $z(x_i) - z(x_i + h)$ ——任意两测站间相距 h 的降水差值。

由于方向原因, 在相同间距内可能存在多个半变异值, 因此可按方向对半变异图进行观察, 若空间相关性具有方向性, 则其方向上的半变异将比其他方向上变化得更快, 该现象也称之为各向异性; 与此相反, 则为各向同性, 在各向同性中空间相关性不随方向而仅随距离变化^[14]。

对降水数据按步长(距离)分组, 选择步长值和步长数的一般规则为: 步长值乘以步长组数小于或等于样本点间最大距离的一半^[14]。经多次试调, 设定步长为 50 km, 步长组为 10, 搜索带宽为 2.5 km 时效果最佳。从图 3 中可以看出, 南北(11°)方向为各向同性, 降水空间相关性不随方向变化, 而仅随距离变化, 东北—西南(44°)方向为各向异性, 即降水空间相关性不仅随距离变化, 而且也随方向发生改变。

3.2 不同模型的半变异拟合

用半变异图测定降水量空间自相关性后, 常需选

取某一种数学函数或模型来拟合半变异, 用于估算任意给定距离的半变异。在已有研究中, 鲁振宇等^[15]对黄河源区 65 个气象站点 1990—2001 年的降水数据进行空间插值、拟合时, 发现指数模型拟合精度较高, 能较为真实地反映该研究区降水的空间分布规律; 江善虎等^[16]利用老哈河流域 52 个气象站点 1994—2005 年的逐日降水数据进行空间插值时发现, 采用球面模型的拟合精度较高, 能较好地反映该流域降水空间的变异性; 秦伟良等^[17]针对扬州市 1999—2008 年 1 月降雨数据进行空间插值分析时发现, 采用高斯模型进行拟合能更真实地反映扬州市 10 a 来降雨空间分布状况; 陈晶晶等^[18]对天津市降水空间数据进行分析时发现, 有理二次曲线模型的精度比其它模型高, 其更能反映天津降水的空间分布规律。为探讨基于不同函数拟合半变异模型对江西省降水量空间插值精度的影响, 本研究分别选取球面、指数、高斯及有理二次曲线模型进行拟合。

由前述东北—西南(44°)搜索方向的半变异云图(图 3)可知, 此方向存在各向异性, 因此在用数学函数拟合时, 需设置各向异性、搜索方向为 44° 。由此获得江西省 30 a 年均降水量的四种数学模型拟合的特征参数(表 1)。由于存在测量误差和空间变异, 两个采样点非常接近时, 它们的半变异函数值不为 0, 即存在块金值, 测量误差和空间变异共同产生了块金值, 而误差随块金值的增大而增大^[19], 对比表 1 中的各项参数可知, 指数模型的块金值最小为 6 553.4 m, 其降水量空间插值误差相应最小; 这四种模型的基台值变化不大, 且基台值本身对插值结果影响较小^[19];

至于变程,存在最佳变程,球面模型变程最小为 468 282 m。综上可初步得出球面、指数模型用于拟

合江西省降水量空间分布更为合理,不同模型拟合效果对插值精度的影响还需通过后续交叉验证来检验。

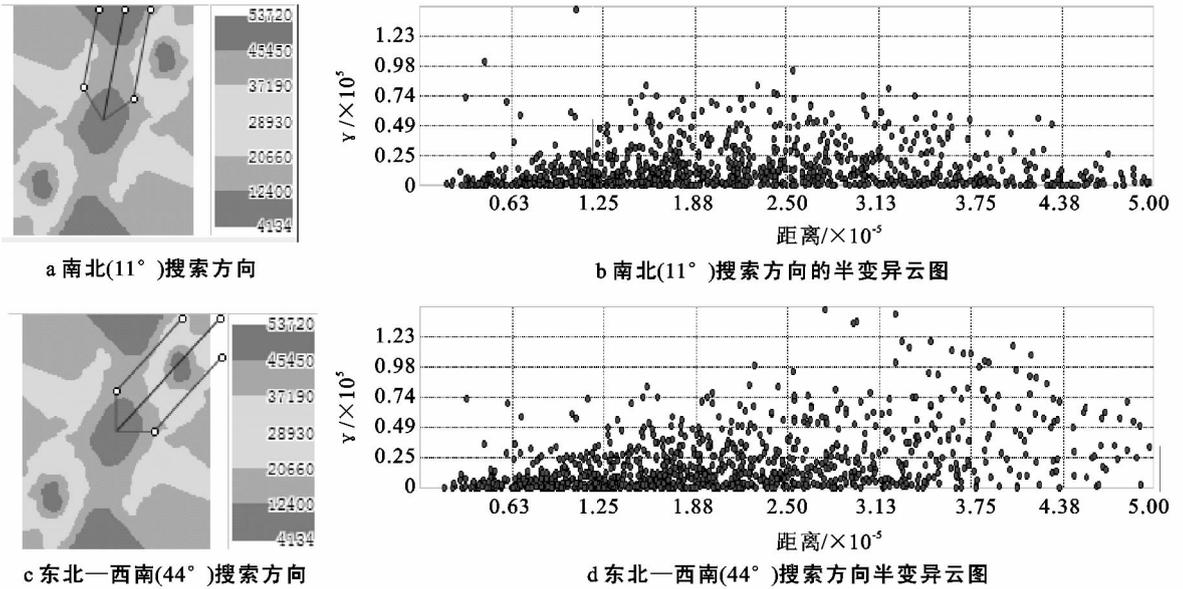


图 3 建模站点各方向半变异云图

表 1 四种半变异函数拟合的特征参数

模型	块金值/m	基台值/m	变程/m
球面	8279.8	26853.8	468169
指数	6553.4	26238.4	496236
高斯	9866.0	27211.0	493282
有理二次曲线	9346.7	24360.7	496236

4 空间插值及其检验

4.1 空间插值

泛克里格插值法(Universal Kriging)是以空间自相关性为基础,利用已有区域点的原始数据与半变异函数的结构性,求取研究区内未知点值的无偏最优估值方法^[20]。该方法假设数据中存在主导趋势,且这种趋势可用确定的函数或多项式来拟合。进行插值时,首先需分析原始数据中存在的变化趋势,确定合适的拟合模型,然后对残差数据(即原始数据与趋势数据之差)进行克里格分析,进而将趋势分析与残差分析的克里格结果求和,获得最终结果。由此可见,空间插值拟合时使用泛克里格方法优于普通克里格^[21]。

分别运用 4 种半变异模型的泛克里格插值方法对江西 30 a 年均降水量进行插值,结果如图 4 所示。分析发现,高斯模型拟合效果相对不理想,譬如万年站点附近降水量等值线数值为 1 750 mm,而事实上万年站点的实测多年平均降水量仅为 1 435 mm,拟合误差相对较大;此外高斯模型在短距离内半变异值的变化程度大于其它 3 种模型,这是由于函数本身所导致的,且等值线较为不平滑,对预测结果影响较大。相比而

言,球面、指数模型拟合效果较为理想,在万年站点附近均形成了降水低谷中心,且绝大部分拟合值与实测值误差较小,符合江西多年平均降水实际情况。

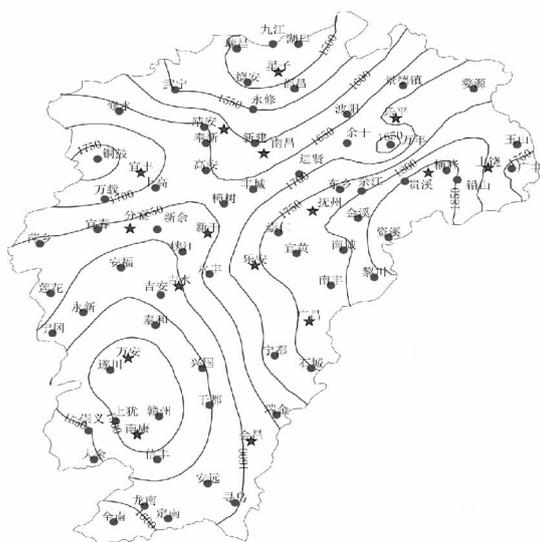
4.2 结果检验

为了验证 4 种函数模型插值结果的精确度和可靠性,需利用一种客观的检验方法进行评价,常见的检验方法有交叉验证法和验证点检验法^[22]。一般符合以下准则的模型为最优:① 标准平均值(Mean Standardized)接近于 0;② 均方根预测误差(Root Mean Square)是衡量预测值与实测值之间的接近程度,应尽可能最小;③ 平均标准误差(Average Mean Error)应尽可能地接近均方根预测误差;④ 标准均方根预测误差(Root Mean Square Standardized)接近于 1^[23]。为此,本文应用交叉验证法和验证点检验法对前述四种模型插值结果进行检验,以确定精确度和可靠性最优的模型(表 2)。

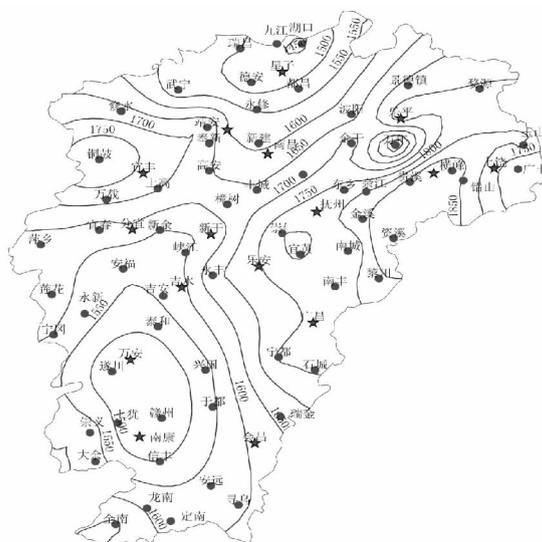
基于前述的四条评价准则,结合建模站点交叉验证结果(表 2)可知,高斯模型均方根预测误差为 110.9、平均标准误差为 107.7,这两者差值较其它三种模型大,标准均方根预测误差为 1.126,也较其它三种模型的误差大,由此可知高斯模型建模站点拟合效果相对较差;而球面模型和指数模型的均方根预测误差分别为 111.9 和 110.8,平均标准误差分别为 110.6 和 110.1,这两者较为接近,标准均方根预测误差分别为 1.024 和 1.023,较为接近 1,表明其误差较小,说明这两种模型的拟合效果较优,其中以指数模型拟合效果最优;高斯模型的数学函数在变程范围内

不能很好地拟合江西省多年平均降水量的空间分布特征,而指数和球面模型正好弥补了这一缺点。此外,综合验证站点检验结果可知,拟合效果较好的同样也是球面模型和指数模型,其标准均方根预测误差分别为 1.105 和 1.104,优于其它两种模型,表明该模

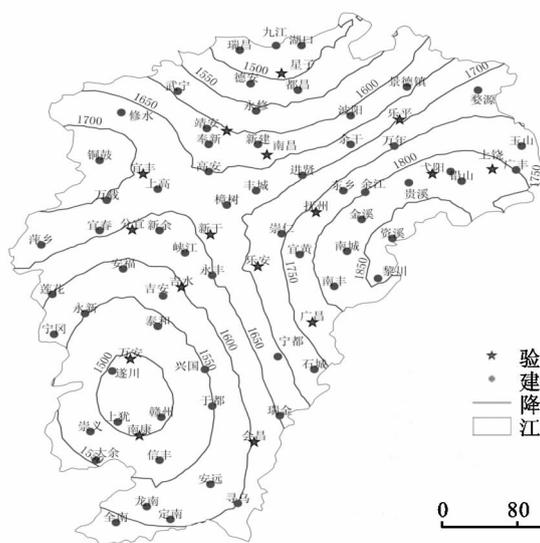
型能较准确地反映江西省多年平均降水量的空间分布特征。综上所述,就模型拟合结果的优劣程度而言,适用于江西省多年平均降水量空间插值拟合模型的优劣顺序依次为:指数模型>球面模型>有理二次曲线模型>高斯模型。



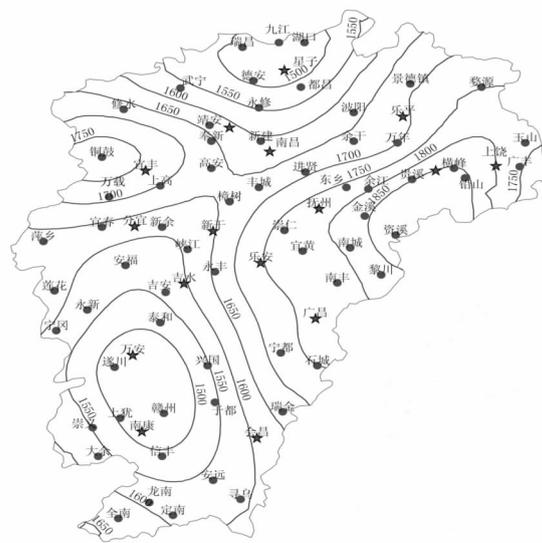
球面模型插值结果



指数模型插值结果



高斯模型插值结果



有理二次曲线插值结果

图 4 4 种半变异模型插值结果

表 2 模型检验结果

模型	验证方法	标准平均值/mm	均方根预测误差/mm	平均标准误差/mm	标准均方根预测误差
球面	建模站点交叉验证	0.0063	111.9	110.6	1.024
	验证站点检验	-0.1251	124.7	111.3	1.105
指数	建模站点交叉验证	0.0027	110.8	110.1	1.023
	验证站点检验	-0.1123	123.2	111.7	1.104
高斯	建模站点交叉验证	0.0108	110.9	107.7	1.126
	验证站点检验	-0.1266	125.1	108.7	1.145
有理二次曲线	建模站点交叉验证	-0.0032	111.9	112.4	1.111
	验证站点检验	-0.1095	122.4	113.3	1.107

5 结论与讨论

选取江西省 81 个完整时间序列的气象站 30 a (1976—2005 年) 年均降水数据, 检验数据分布特征后, 运用泛克里格中不同半变异模型对降水数据进行拟合, 而后通过基于 65 个建模站点的交叉验证法和 16 个验证站点的检验法对空间插值结果进行检验, 得出以下结论:

(1) 江西省降水量总体分布格局呈现为南多北少、东多西少, 30 a 年均降水量的莫兰指数为 0.65, 表明其空间相关性较强; 降水数据基本服从正态分布, 进行泛克里格插值计算时, 无需进行数据变换。

(2) 江西省多年平均降水存在较大的空间变异性, 南北(11°) 方向为各向同性, 东北—西南(44°) 方向为各向异性, 因此在使用不同半变异函数模型进行插值时, 需考虑方向因素。

(3) 采用泛克里格的指数模型和球面模型插值法能较为准确地反映江西省多年平均降水量空间分布状况, 其中以指数模型的拟合效果最优, 以高斯模型的拟合效果最劣。

然而, 本文中泛克里格插值中的指数模型与球面模型插值法应用于江西省宏观尺度降水空间拟合的精度虽较高, 但影响江西省降水量空间分布的因素较多, 如地形、高程、纬度、气候条件和下垫面状况等, 此外, 不同时空尺度下所选择的插值模型也可能不尽相同, 故在江西省降水空间插值的微观尺度拟合时需综合考虑以上各因素, 该问题可在后续研究中继续深入。

参考文献:

- [1] 王常森, 陶月赞, 方必和. 淮北平原年降水量空间插值模型的比选[J]. 水文, 2012, 32(2): 49-53.
- [2] 段文明, 穆兴民, 高鹏, 等. 嘉陵江流域近 50 年降水量演变规律分析[J]. 水土保持研究, 2012, 19(1): 97-101.
- [3] 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): 811-817.
- [4] 王舒, 严登华, 秦天玲, 等. 基于 PER-Kriging 插值方法的降水空间展布[J]. 水科学进展, 2011, 22(6): 756-762.
- [5] 虎雄岗, 谈树成, 金艳珠, 等. 基于地形高程的云南省降水量空间插值方法研究[J]. 热带地理, 2012, 32(4): 370-377.
- [6] 汪翠英, 杨勤科, 郭伟玲, 等. 区域水土流失研究中降雨量空间插值方法选取[J]. 水土保持研究, 2008, 15(2): 88-91.
- [7] 赵登忠, 张万昌, 宣勇. 基于小波分析的日降水时空模型构建及模拟[J]. 人民长江, 2011, 42(23): 19-22.
- [8] 胡庆芳, 杨大文, 王银堂, 等. 利用全局与局部相关函数分析流域降水空间变异性[J]. 清华大学学报: 自然科学版, 2012, 52(6): 778-784.
- [9] 朱会义, 贾绍凤. 降雨信息空间插值的不确定性分析[J]. 地理科学进展, 2004, 3(1): 34-41.
- [10] Lawrence E B, David L P, Steven W R. Forest ecosystem processes at the watershed scale: basis for distributed simulation[J]. Ecol. Model, 1991, 5(6): 171-196.
- [11] Wong K W, Wong P M, Gedeon T D, et al. Rainfall prediction model using soft computing technique[J]. Soft Comput, 2003, 7(6): 434-438.
- [12] 杨爱萍, 杜筱玲. 江西省降水集中程度的变化特征[J]. 中国农业气象, 2010, 31(4): 512-516.
- [13] 于晓艳, 马劲松, 朱敬芳, 等. 基于地统计学的江西省年降水量插值研究[J]. 测绘科学, 2011, 36(4): 83-85.
- [14] Kang-tsung Chang, 陈建飞. 地理信息系统导论[M]. 北京: 清华大学出版社, 2008.
- [15] 鲁振宇, 杨太保, 郭万钦. 降水空间插值方法应用研究: 以黄河源区为例[J]. 兰州大学学报: 自然科学版, 2006, 42(6): 11-14.
- [16] 江善虎, 任立良, 雍斌, 等. 老哈河流域降水的空间插值方法比较[J]. 干旱区资源与环境, 2010, 24(1): 80-84.
- [17] 秦伟良, 刘悦. 空间插值法在降水分布中的应用[J]. 南京信息工程大学学报: 自然科学版, 2010, 2(2): 162-165.
- [18] 陈晶晶, 胡蓓蓓, 王军, 等. 天津降水数据的空间插值分析[J]. 安徽师范大学学报: 自然科学版, 2010, 33(4): 382-387.
- [19] 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.
- [20] 张慧智, 史学正, 于东升, 等. 中国土壤温度的空间插值方法比较[J]. 地理研究, 2008, 27(6): 1299-1306.
- [21] 牛文杰, 朱大培, 陈其明. 泛克里金插值法的研究[J]. 计算机工程与应用, 2001, 13(1): 73-75.
- [22] 孟庆香, 刘国彬, 杨勤科. 基于 GIS 的黄土高原气象要素空间插值方法[J]. 水土保持研究, 2010, 17(1): 10-14.
- [23] 刘爱丽, 王培法, 丁圆圆. 地统计学概论[M]. 北京: 科学出版社, 2012.