

Comparison and Analysis of Different Variogram Functions Models in Kriging Interpolation of Daily Rainfall

Qingjing Wang¹, Chongyu Xu^{1,2}, Hua Chen¹

¹State Key Laboratory of Water Resources and Engineering Science, Wuhan University, Wuhan Hubei

²Department of Geosciences, University of Oslo, Oslo Norway

Email: qjing.wang@whu.edu.cn

Received: Oct. 4th, 2016; accepted: Oct. 23rd, 2016; published: Oct. 26th, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Ordinary Kriging is a wide-used method of geostatistical interpolation. However, it is only one variogram function that is often used to calculate the estimated value depending on one's experience. Meanwhile, there is lesser study about daily rainfall comparison of different variogram functions in ordinary Kriging because most of the recent researches emphasize yearly and monthly rainfall. To obtain the better fitted variogram function in ordinary Kriging for daily rainfall interpolation within four different variogram functions (exponential, spherical, gauss, linear). The study leans on 16-year (1990-2005) daily rainfall data at 43 raingages in and around Mishui Basin. According to the comparison of three aspects including the correlation coefficients, examine index, accuracy of different precipitation grade, the result of cross-validation shows that: 1) exponential and spherical functions can fit the daily rainfall interpolation better in ordinary Kriging; 2) the estimated value in ordinary Kriging is generally smaller than observed value on days of heavy rainfall; 3) the error of daily rainfall interpolation increases with the precipitation grade.

Keywords

Ordinary Kriging, Variogram Function, Daily Rainfall, Spatial Interpolation, Cross-Validation

克里金日降水插值的不同变异函数比较分析

汪青静¹, 许崇育^{1,2}, 陈 华¹

作者简介: 汪青静(1993-), 女, 湖北咸宁人, 硕士研究生, 主要从事水文预报方面研究。

¹武汉大学水资源与水电工程科学国家重点实验室, 湖北 武汉

²挪威奥斯陆大学地学系, 挪威 奥斯陆

Email: qjing.wang@whu.edu.cn

收稿日期: 2016年10月4日; 录用日期: 2016年10月23日; 发布日期: 2016年10月26日

摘要

普通克里金法为一种广泛使用的地理统计的插值方法,但一般都只依据经验使用一个变异函数来计算插值结果。同时,现有研究着重研究年、月尺度下普通克里金法不同变异函数的比较,对日降水尺度研究较少。为了得到普通克里金法在日降水插值适应性较好的变异函数,比较和分析普通克里金法四种常用不同变异函数(指数、球型、高斯、线性)。选取了湘江涿水所在子流域及子流域周边一共43个雨量站1990~2005年逐日降水资料插值,从相关系数、检验指标、不同降水级别准确率这三方面比较插值结果,交叉验证的结果表明:1) 指数和球型函数对普通克里金法的日降水插值适应性较好;2) 普通克里金法的插值结果在降水量大的日子普遍偏小;3) 随着降水量级的增大,日降水插值的误差明显增大。

关键词

普通克里金, 变异函数, 日降水, 空间插值, 交叉验证

1. 引言

降水观测值对于气象学、气候学、水文学、地质分析、环境监控都必不可少。降水资料通常是通过已有的雨量站、气象站(包括降水、温度和风速等)收集,因此这些数据都是离散的、局部的和有限的空间点。然而,对于单独的降水站点资料作为径流估计的输入会带来很大的不确定性[1] [2]。时空连续的降水资料,需要根据已知降水站点的资料通过空间插值得到。

现有的空间插值方法可以被概括为确定性和地理统计的插值方法[3],地理统计的方法正在广泛的使用。地理统计提供了一种考虑空间变量的插值方法,经常使用变异函数来描述空间变量结构。克里金法提供了一个在有限区域内对空间变量进行无偏最优估计的方法。Borga 和 Vizzaccaro [4]用线性函数作为普通克里金插值的变异函数和复二次函数曲面拟合的插值方法进行降水插值比较,发现普通克里金法的插值精度较好。Chen 等[5]用线性函数作为普通克里金插值的变异函数,比较了普通克里金法、最邻近法、局部多项式法、径向基法、倒距离权重法在中国 1951~2005 年的日降水插值结果,得出普通克里金法的精度最高。现有关于降水的空间插值研究中,大多数研究倾向于对不同插值方法进行比较,很少涉及到插值方法的参数设置和变异函数的选择。许多研究在应用克里金法时多是根据经验人为的选定一个变异函数,因而缺乏可靠的依据。S.Ly 等[6]分析了在比利时乌尔特河流域不同变异函数对日降水插值影响,发现最合适的变异函数是高斯函数。Lin 和 Chen [7]认为变异函数的计算对于降水插值精度有很大影响,变异函数没有拟合好,克里金法将不会产生最好的插值。因此,变异函数的选择对于克里金插值的方法至关重要。

由于降水具有间断性和空间不连续性,而且日降水零值较多,因此降水的空间插值与其他要素相比具有更大的困难[8]。目前,较多研究者考虑的是年降水、月降水尺度下的克里金插值变异函数的比较,张余庆等[9]研究了江西多年平均降水量空间插值模型的选取与比较,王常森等[10]研究了淮北平原年降水量空间插值模型的比选。日降水的变异函数比较研究较少,但随着分布式水文模型的发展,对日降水插值的要求也越来越高。

本文的目的是得到普通克里金日降水插值适应性较好的变异函数。选取了普通克里金插值常用的 4 种变异

函数, 设置变异函数参数后计算得估计值, 使用交叉验证的思路, 对插值结果使用不同的评判方法, 比较分析 4 种变异函数日降水的插值结果, 得到适应性较好的变异函数。

2. 研究方法

2.1. 插值方法

普通克里金法是一种经常使用的地理统计的插值方法, 而地理统计是一种既考虑样本值又重视样本空间位置及样本之间相关性的方法, 通常使用变异函数作为描述空间相关性的核心。

2.1.1. 变异函数

变异函数用来刻画区域化变量在空间中的相关程度, 定义为区域化变量在相距为 h 的任意两点处的平方均值的一半, 其计算公式如下:

$$\bar{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z(x_i) - Z(x_i + h))^2 \quad (1)$$

式中: $\bar{\gamma}(h)$ 为该点处变异函数的值; $N(h)$ 为分离距离为 h 的采样点对的数目; $Z(x_i)$ 和 $Z(x_i + h)$ 为采样间隔为 h 的已知点的降水数据。

根据变异函数的特性, 选取适当的半变异理论函数, 本文选取的半变异理论函数有常用的四个, 分别为球面函数、高斯函数、指数函数、线性函数, 其公式如下。

球型函数(Spherical function):

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 + C \left(\frac{3h}{2\alpha} - \frac{1}{2} \frac{h^3}{\alpha^3} \right) & 0 < h \leq \alpha \\ C_0 + C & h > \alpha \end{cases} \quad (2)$$

高斯函数(Gauss function):

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 + C \left(1 - e^{-\frac{3h^2}{\alpha^2}} \right) & h > 0 \end{cases} \quad (3)$$

指数函数(Exponential function):

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 + C \left(1 - e^{-\frac{3h}{\alpha}} \right) & h > 0 \end{cases} \quad (4)$$

线性函数(Linearfunction):

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 + C \left(\frac{h}{\alpha} \right) & 0 < h \leq \alpha \\ C_0 + C & h > \alpha \end{cases} \quad (5)$$

在以上四个函数中, h 为分离距离, C_0 表示块金值, $C_0 + C$ 表示基台值, α 表示变程, 即曲线到达基台值时所对应的距离。

在选择分离距离 h 时, 要考虑到在大分离距离下的测量资料代表样本采集地边缘的方差结构, 而不是样本主流的方差结构, 近距离的变异函数值比远距离的变异函数值所起作用更大。因此要将分离距离控制在有意义的研究范围内, 通常应保证分离距离 $|h| \leq L/2$ [11], L 是研究区域沿某方向的最大距离。

受样本数量限制, 步长值的选取不宜过小。在每一个分离距离上用来计算样本变异函数的数值一般应大于 30 个点对[11], 点对数太小的变异函数值不可用。因此虽然适当减小步长值一般能提高函数拟合精度, 但如果参与计算的数据点对太少, 则只能增加步长值。

2.1.2. 普通克里金法

普通克里金(Ordinary Kriging)提供了一个在有限区域内对空间变量进行无偏最优估计的方法, 是根据样本空间位置不同、样本间相关程度不同, 对每个样品赋予了不同的权, 进行滑动加权平均, 以估计待测点的值。

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (6)$$

式中: $Z^*(x_0)$ 为待测点的估算值; $Z(x_i)$ 为第 i 个样本点的实测值, n 为参与计算的实测样本个数, λ_i 为第 i 个样本点的权重系数。

而权重是根据克里金插值的无偏估计和方差的最小得到。公式如下:

$$\begin{cases} \sum_{i=1}^n \bar{\gamma}(x_i - x_j) \lambda_i + \mu = \bar{\gamma}(x_0 - x_j) & (j=1, 2, \dots, n) \\ \sum_{i=1}^n \lambda_i = 1 \end{cases} \quad (7)$$

式中: $\bar{\gamma}(x_i - x_j)$ 为第 i 个和第 j 个样本点之间的半方差, $\bar{\gamma}(x_0 - x_j)$ 为第 j 个样本点与待插点之间的半方差, 这两个量均可以从已拟合的变异函数得到, μ 为拉格朗日乘子, n 为参与计算的实测样本点个数。

2.2. 评判方法

交叉验证

交叉验证法(Cross-validation)通过求出检验站点的实测值与估计值之间的误差, 作为评价插值方法效果的依据。

本文选用平均误差 ME (Mean Error), 用来评估插值方法的无偏性, 其值越小越好; 平均绝对误差 MAE (Mean Absolute Error), 表示插值结果存在系统误差的大小, 其值越小越好; 均方根误差 RMSE (Root Mean Squared Error) 表示插值结果偏离实际值的大小, 其值越小越好。

$$ME = \sum_{i=1}^n \frac{Z^*(x_0) - Z(x_i)}{n} \quad (8)$$

$$MAE = \sum_{i=1}^n \frac{|Z^*(x_0) - Z(x_i)|}{n} \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Z^*(x_0) - Z(x_i))^2}{n}} \quad (10)$$

式中: $Z^*(x_0)$ 为待插点的估算值; $Z(x_i)$ 为第 i 个样本点的实测值, n 为参与计算的实测样本个数。

3. 研究区域及资料

沅水是湘江的一级支流, 本文研究的是沅水所在子流域。该子流域位于北纬 $26^{\circ}1' \sim 27^{\circ}23'$, 东经 $112^{\circ}57' \sim 114^{\circ}6'$ 之间, 整个子流域面积是 9929 km^2 。该子流域高程范围从 $6 \sim 2061 \text{ m}$ (图 1), 东南为山地, 高程超过 1000 m , 其余大都为起伏不平的丘陵与沿河平原。东南部山地主要是变质岩, 中部西南部主要为红岩和第四纪松散堆积物。该子流域是亚热带季风气候, 年均气温在 17°C 以上, 年均降水为 1600 mm , 东南降水较大, 可达 1900 mm 以上, 西南降水较小, 在 1400 mm 左右, 降水从东南向西北递减。全年降水多集中于 $3 \sim 7$ 月, 汛期($4 \sim 9$ 月)降水占全年的 70% 左右。

本文的降水数据来源于沅水所在子流域及子流域周边一共 43 个雨量站(图 1)。这些站点具备 $1990 \sim 2005$ 年完整的日降水数据, 且空间分布比较均匀。以 40 个雨量站作为计算站点, 以悟田洲站、承坪站、打鸟坳站这 3 个站点作为交叉验证的检验站点。

4. 结果与讨论

4.1. 变异函数参数的选择

在计算变异函数参数时, 使用的是沅水所在子流域及周边 40 个计算站点, 其中东江站和清水江站距离最远, 为 $172,561 \text{ m}$, 故在进行半方差分析时, 选取此距离值的一半(约 $86,280 \text{ m}$)作为半方差分析的最大步长值。基于以上分析, 在对现有数据进行分析时, 经过多次步长设置的比较, 最终选取步长值为 $10,785 \text{ m}$, 以保证参与计算半方差的点数至少为 30 , 以此获得的步长组数为 8 。

由于变异函数的使用, 克里金插值会产生不好的结果——插值结果小于零。通常, 这里有两种方法来避免不好的插值结果: 一是通过 Deutsh [12] 给出的经验纠正; 二是直接将小于零的值替代为零。本文是直接小于零的值替代为零。

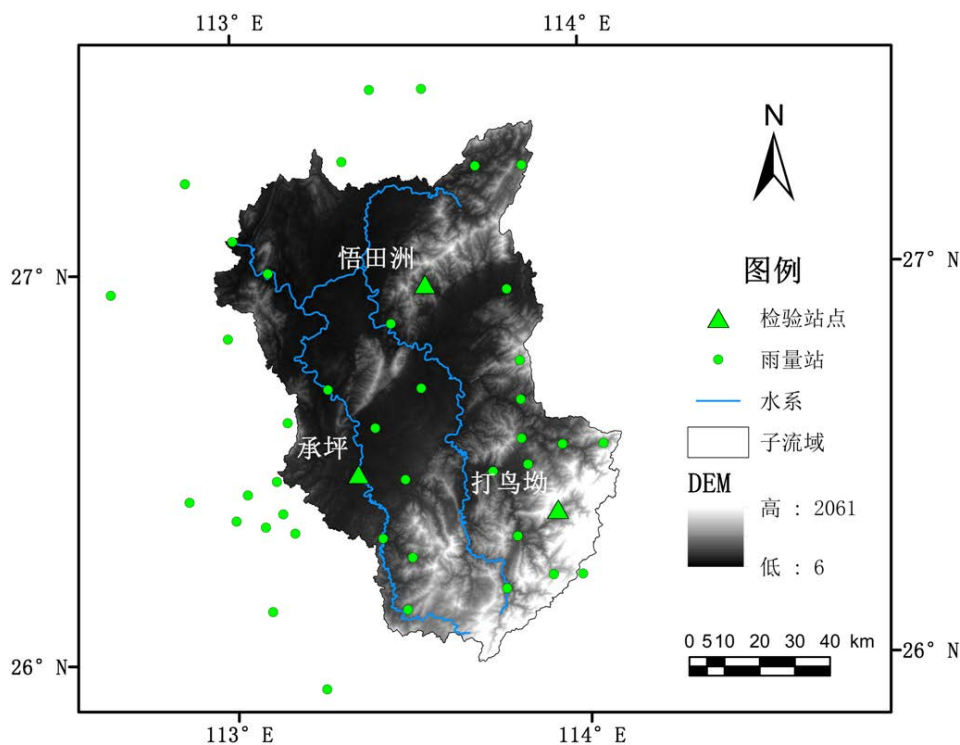


Figure 1. DEM and distribution of rainfall stations in Mishui Basin
图 1. 沅水子流域数字高程及站点水系分布图

4.2. 交叉验证结果与分析

在对普通克里金插值结果进行评估时，对于所有站点日降水都为零的日子，我们认为在整个流域都是没有降水的，因此不会对这些日子进行插值。本文已有的 16 年日降水资料中，无雨的日子有 983 天。

4.2.1. 相关系数比较

图 2 给出普通克里金法三个检验站点的 4 种不同变异函数 16 年日降水量实测值和估计值结果的对比及其相关系数。从相关系数可以看到，这 4 种不同变异函数的实测值与估计值均有较高相关性，相关系数在 0.8 左右。由于所选检验站点的位置不同、所使用插值的已知站点不同，这三个检验站点的相关系数不尽相同，但是对于每一个检验站点的 4 种不同变异函数相关性的比较，可以看到指数函数的相关性最高，球型函数的相关性次之。其中指数函数和球型函数比较相近，高斯函数和线性函数相关性比较相近。从图中也可以看到，对于每一个检验站点，指数函数和球型函数的估计值在斜率为 1 的斜线附近较为集中，而高斯函数和线性函数的估计值在斜

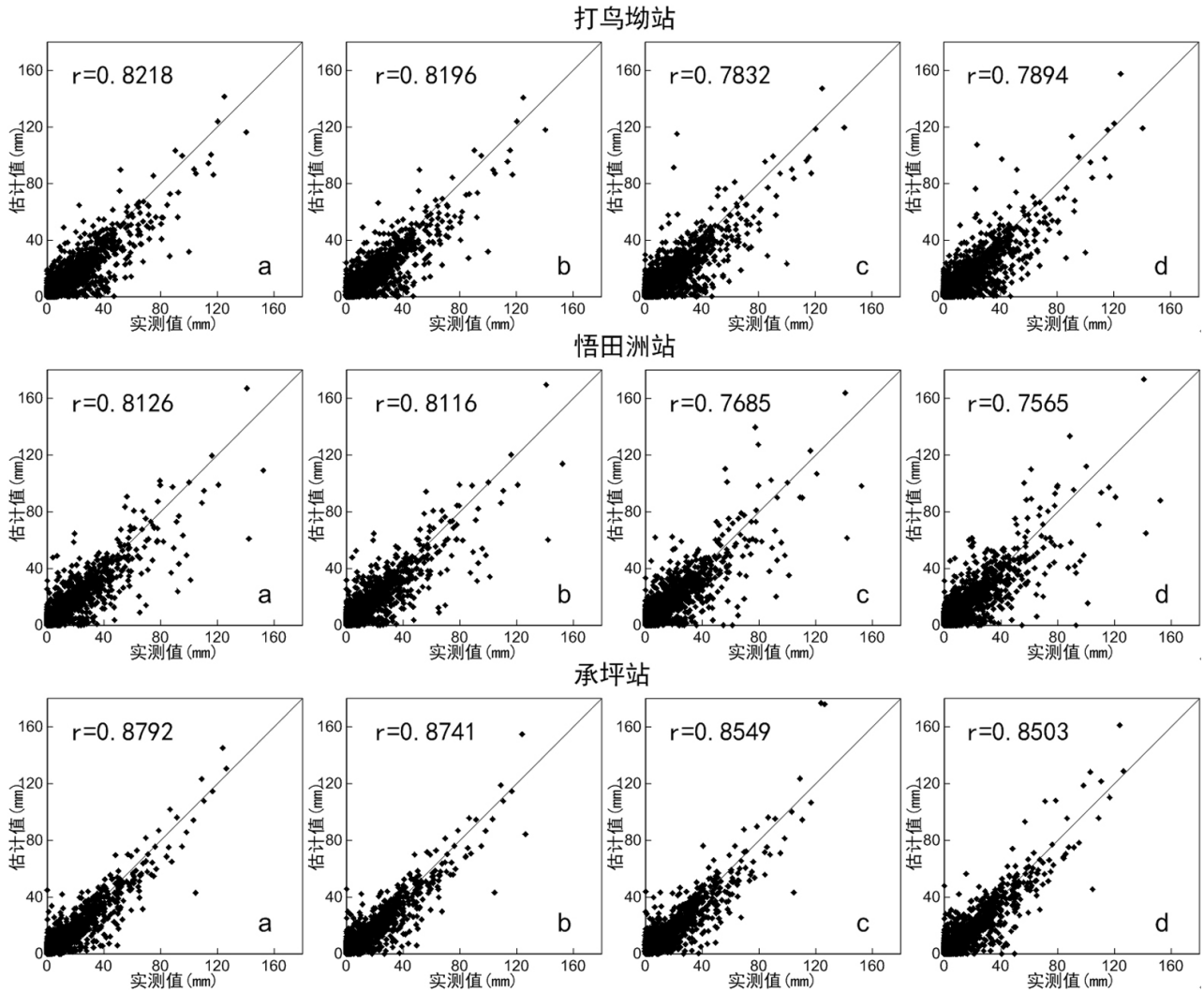


Figure 2. The correlation coefficients and comparison of four different variogram functions between estimated and observed value for daily rainfall (a) Exponential function; (b) Spherical function; (c) Gauss function; (d) Linear function

图 2. 四种不同变异函数日降水实测值与估计值对比图及其相关系数 (a)指数函数; (b)球型函数; (c)高斯函数; (d)线性函数

率为 1 的斜线附近较为分散, 偏离较大, 导致相关系数偏低, 可以认为使用高斯函数和线性函数插值出现极值误差的几率较大。

4.2.2. 不同检验指标比较

本文采用平均误差 ME、平均绝对误差 MAE 和均方根误差 RMSE 3 个指标来衡量 4 种不同变异函数的插值效果, 其结果如表 1。从这三个检验站点的 ME 来看, 高斯函数的误差最大, 明显高于指数函数、球型函数、线性函数。从 MAE 和 RMSE 来看, 指数函数的误差值最小, 球型函数的误差次之, 但是指数函数和球型函数的误差明显的小于高斯函数和线性函数, 说明指数函数的估计值误差范围较小, 与实际值更为接近, 能更好的适应日降水插值。

由于该地区无降水的日子有 983 天, 将会影响到对较大雨量情况的检验结果, 因此对降水量大于等于 10 mm 的情况, 也进行了检验, 结果如表 2。表 2 中三个检验站点的 ME 均小于零, 说明这 4 种不同变异函数的估计值相较于实测值系统的偏小, 相比于表 1 的 ME, 可以得到降水量大的日子普通克里金插值结果具有一定的减小作用。同时, 表 2 的 MAE 和 RMSE 相较于表 1 均明显增大了很多, 说明随着降水量的增大, 插值的误差范围也会增大。但随着降水量的增大, 指数函数和球型函数的误差范围非常接近, 指数函数的误差不再是最小的了, 但还是明显的低于高斯函数和线性函数。

4.2.3. 不同降水级别的插值准确率比较

以降水量(R) 10, 25, 50 mm 为标准, 对大于等于界限的插值结果进行分析。如果实测值大于等于标准值, 其插值结果也大于等于标准值, 则认为插值准确; 如果实测值大于等于标准值, 插值结果小于标准值, 则认为插值结果不准确。分别统计满足这 4 种不同变异函数情况下的样本数, 然后除以二者的样本数总和, 可以得到这 4 种变异函数所占的比例。

从图 3 三个检验站点的准确率来看, 随着降水量级的增大, 指数函数误差最小的优势不再明显, 但无论在哪个降水量级, 指数函数和球型函数插值结果的准确率非常接近, 并且高于高斯函数和线性函数。随着降水量级的增大, 插值的准确率有明显较小的趋势。在 $R \geq 10$ mm (中雨) 时, 这三个检验站点的准确率在 80% 左右; 在 $R \geq 25$ mm (大雨) 时, 准确率在 70% 左右; 在 $R \geq 50$ mm (暴雨) 时, 准确率在 60% 左右。由此可见, 在降水量较大的情况, 尤其是对暴雨的插值, 普通克里金插值的方法不是很好。

Table 1. The cross-validation results of four different variogram functions between estimated and observed value
表 1. 4 种不同变异函数交叉验证检验站点结果

站点	平均误差 ME/mm				平均绝对误差 MAE/mm				均方根误差 RMSE/mm			
	指数	球型	高斯	线性	指数	球型	高斯	线性	指数	球型	高斯	线性
打鸟坳站	-0.35	-0.35	-0.39	-0.23	1.95	1.96	2.11	2.15	25.58	25.85	31.26	39.74
悟田洲站	0.11	0.14	0.30	0.13	1.85	1.86	2.15	2.10	24.35	24.54	34.87	31.94
承坪站	0.08	0.08	0.17	0.15	1.43	1.45	1.58	1.60	13.76	14.34	16.58	17.18

Table 2. The cross-validation results of four different functions between estimated and observed values (daily rainfall ≥ 10 mm)
表 2. 四种不同变异函数交叉验证检验站点结果(日降水量 ≥ 10 mm)

站点	平均误差 ME/mm				平均绝对误差 MAE/mm				均方根误差 RMSE/mm			
	指数	球型	高斯	线性	指数	球型	高斯	线性	指数	球型	高斯	线性
打鸟坳站	-4.86	-4.85	-5.06	-4.46	7.85	7.85	8.56	8.69	141.06	141.80	174.84	230.47
悟田洲站	-2.73	-2.60	-2.55	-2.82	6.46	6.46	7.39	7.63	123.38	123.01	181.73	169.12
承坪站	-2.40	-2.43	-2.47	-2.36	4.74	4.82	5.07	5.28	63.46	66.34	76.56	79.68

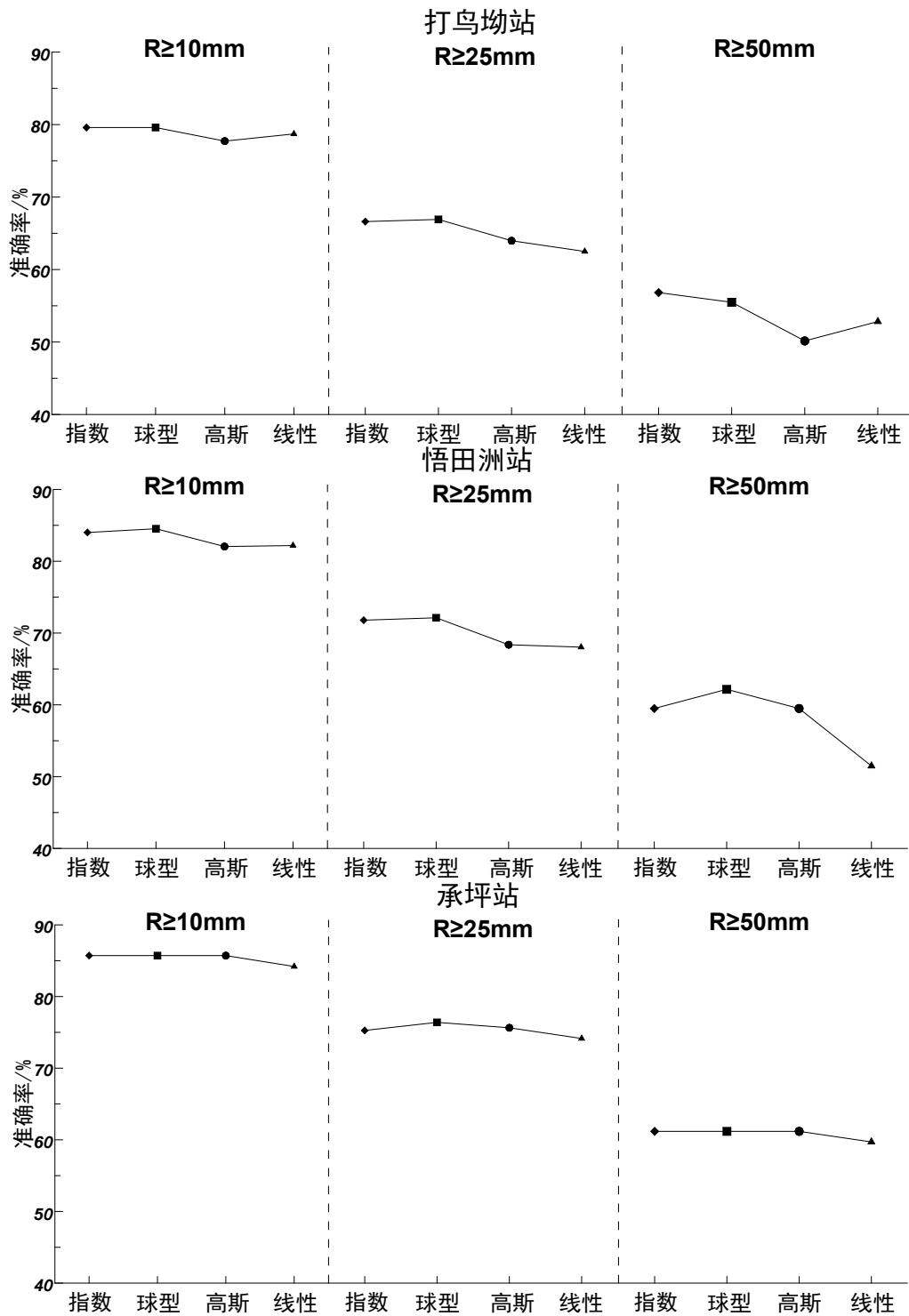


Figure 3. Accuracy of different precipitation grade of four different variogram functions

图 3. 四种不同变异函数在不同降水级别下的准确率

5. 结论

本文研究了普通克里金法在 4 种不同变异函数的情况下, 对湘江沅水子流域 40 个雨量站的日降水资料插值,

设置变异函数步长值为 10,785 m, 步长组数为 8, 从相关系数、检验指标、不同降水级别准确率这三个方面对交叉验证的 3 个检验站点的插值结果进行比较分析, 得出以下结论:

- 1) 指数和球型函数对于普通克里金日降水插值的适应性较好。指数函数在日降水量级较小时适应性更好; 在日降水量级增大时, 指数和球型函数对普通克里金插值的适应性均较好。
- 2) 普通克里金法的插值结果在降水量大的日子普遍偏小。
- 3) 随着降水量级的增大, 日降水的误差明显增大。

基金项目

国家自然科学基金重点项目(51539009)。

参考文献 (References)

- [1] FAURÈS, J. M., GOODRICH, D. C., WOOLHISER, D. A., et al. Impact of small-scale spatial variability on runoff modeling. *Journal of Hydrology*, 1995, 173(1): 309-326. [http://dx.doi.org/10.1016/0022-1694\(95\)02704-S](http://dx.doi.org/10.1016/0022-1694(95)02704-S)
- [2] CHAUBEY, I., HAAN, C. T., GRUNWALD, S., et al. Uncertainty in the model parameters due to spatial variability of rainfall. *Journal of Hydrology*, 1999, 220(1-2): 48-61. [http://dx.doi.org/10.1016/S0022-1694\(99\)00063-3](http://dx.doi.org/10.1016/S0022-1694(99)00063-3)
- [3] 易湘生, 李国胜, 尹衍雨, 彭景涛. 土壤厚度的空间插值方法比较——以青海三江源地区为例[J]. *地理研究*, 2012, 31(10): 1793-1805.
YI Xiangsheng, LI Guosheng, YIN Yanyu and PENG Jingtao. Comparison on soil depth prediction among different spatial interpolation methods: A case study in the Three-River Headwaters Region of Qinghai Province. *Geographical Research*, 2012, 31(10): 1793-1805. (in Chinese)
- [4] BORGA, M., VIZZACCARO, A. On the interpolation of hydrologic variables: Formal equivalence of multiquadratic surface fitting and kriging. *Journal of Hydrology*, 1997, 195(1-4): 160-171. [http://dx.doi.org/10.1016/S0022-1694\(96\)03250-7](http://dx.doi.org/10.1016/S0022-1694(96)03250-7)
- [5] CHEN, D., OU, T., GONG, L., et al. Spatial interpolation of daily precipitation in China: 1951-2005. *Advances in Atmospheric Sciences*, 2010, 27(6): 1221-1232. <http://dx.doi.org/10.1007/s00376-010-9151-y>
- [6] LY, S., CHARLES, C. and DEGRÉ, A. Geostatistical interpolation of daily rainfall at catchment scale: The use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrology & Earth System Sciences*, 2011, 15(7): 2259-2274. <http://dx.doi.org/10.5194/hess-15-2259-2011>
- [7] LIN, G. F., CHEN, L. H. A spatial interpolation method based on radial basis function networks incorporating a semivariogram model. *Journal of Hydrology*, 2004, 288(288): 288-298. <http://dx.doi.org/10.1016/j.jhydrol.2003.10.008>
- [8] 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. *Journal of Hydrology*, 2012, 208(3): 187-193.
- [9] 张余庆, 陈昌春, 尹义星, 杨绪红. 江西多年平均降水量空间插值模型的选取与比较[J]. *水土保持研究*, 2013, 20(4): 69-74.
ZHANG Yuqing, CHEN Changchun, YIN Yixing and YANG Xuhong. Spatial interpolation model selection of multi-year average precipitation in Jiangxi Province. *Research of Soil and Water Conservation*, 2013, 20(4): 69-74. (in Chinese)
- [10] 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.
WANG Changsen, TAO Yuezan and FANG Bihe. Comparison of spatial interpolation models for annual precipitation in Huaibei Plain. *Journal of China Hydrology*, 2012, 32(2): 49-53. (in Chinese)
- [11] 张康聪. 地理信息系统导论[M]. 电子工业出版社, 2014.
CHANG K.-T. Introduction to geographic information systems. Electronic Industry Publishing House, 2014.
- [12] DEUTSCH, C. V. Correcting for negative weights in ordinary kriging. *Computers & Geosciences*, 1996, 22(7): 765-773. [http://dx.doi.org/10.1016/0098-3004\(96\)00005-2](http://dx.doi.org/10.1016/0098-3004(96)00005-2)