

## 考虑数据变换的泾河流域月降雨空间插值\*

刘 鹏<sup>1</sup>, 张万昌<sup>2</sup>

(1. 南京大学 国际地球系统科学研究所, 南京 210093; 2. 中国科学院 大气物理所东亚区域气候-环境重点实验室, 全球变化东亚区域研究中心, 北京 100029)

**摘 要:** 对泾河流域各站点 1951–2005 年多年平均月降雨量数据进行偏斜度分析, 并分别做平方根、立方根、对数变换; 根据泾河流域自然地理特征选择普通克里格法和梯度距离平方反比法对变换前后的月均降雨量数据进行空间插值, 通过交叉验证的检验方法得到以下结论: 偏斜度分析对于空间插值意义不大; 两种插值方法都是对数变换后的数据插值结果较优, 同时考虑高程的梯度距离平方反比法无论数据变换与否其插值精度都没有普通克里格法的精度高。选择对数变换和普通克里格法进行泾河流域多年平均月降雨量插值效果较好, 为日降水插值的获得提供了可靠的本底文件。

**关键词:** 偏斜度分析; 数据变换; 普通克里格法; 梯度距离平方反比法

中图分类号: P426.614

文献标识码: A

文章编号: 1005-3409(2008)04-0001-04

## The Month Precipitation in Jinghe River Basin Considering Data transformation

LIU Peng<sup>1</sup>, ZHANG Wanchang<sup>2</sup>

(1. International Institute for Earth System Science, Nanjing University, Nanjing 210093, China; 2. START Regional Center for Temperate East Asia, Institute of Atmospheric Physics, CAS, Beijing 100029, China)

**Abstract:** The skewness, square root transformation, cube root transformation and logarithm transformation of the monthly precipitation which was from the stations throughout Jinghe river basin from 1951 to 2005 was analyzed firstly. Then, the spatial interpolation of monthly precipitation before and after data transformation was calculated by Ordinary Kriging and Gradient plus reverse distance squared which were chosen according to the geographical characteristics of the Jinghe River Basin. The cross validation shows that the skewness was not meaningful for spatial interpolation. The logarithm transformation had the best results whether the spatial interpolation was calculated by Ordinary Kriging or Gradient plus reverse distance squared. The results of Ordinary Kriging were better than that of Gradient plus reverse distance squared whether the monthly precipitation was transformed or not. The Ordinary Kriging with logarithm transformation got the best results for daily precipitation spatial interpolation.

**Key words:** skewness analysis; data transformation; Ordinary Kriging; Gradient plus reverse distance squared (GRDS)

### 1 引言

分布式水文模拟考虑了流域自然地理要素的空间分异特性, 可以更深入地揭示水循环的物理过程, 并能及时地模拟出人类活动或下垫面因素的变化对流域水循环过程的影响。降雨量是其重要的输入数据。由于分布式水文模拟考虑了自然地理分异特性, 实践中要求输入的是流域每个栅格位置的降雨量, 而降雨量数据来源于空间分布一般不均匀的气象或者水文台站, 即降雨量数据一般是基于台站的点雨

量, 如何将点雨量数据按照一定的数学方法结合流域地理特征展布到流域每一个栅格上, 并能保证精度一直以来都是相关领域的研究重点<sup>[1-3]</sup>。对流域进行日降雨径流分布式模拟的实践中, 需要输入的则是日降雨插值数据。可以通过把多年平均月降雨插值结果当作本底文件, 然后考虑日降水概率得到精度相对较高的日降雨插值(如 PRISM 插值), 即多年平均月降雨插值在一定程度上影响日降雨插值的精度<sup>[4-5]</sup>。

插值结果主要受插值方法和数据的影响。插值方法的选择要结合特定研究区。泾河是渭河的一级支流, 黄河的二

\* 收稿日期: 2007-05-15

基金项目: 中科院“百人计划”择优支持项目(8-057493); 国家重点基础研究发展规划 973 项目(2006CB400502, 2001CB309404)

作者简介: 刘鹏(1984-), 男, 山东省临沂市人, 硕士研究生, 主要从事遥感与地理信息系统及其在水文学中的应用研究。E-mail: liupeng\_nj@126.com

通信作者: 张万昌(1966-), 男, 研究员, 博导, 研究方向为遥感与 GIS 集成技术, GIS 水文学及水资源。E-mail: zhangwc@tea.ac.cn

级支流,流域面积 45 421 km<sup>2</sup>, 全长 455. 1 km, 流域范围跨 106° 20′ - 108° 48′ E, 34° 24′ - 37° 20′ N, 处于六盘山和子午岭之间。发源于宁夏回族自治区泾源县关山东麓, 由西北向东南流经宁夏、甘肃、陕西三省区的固原、平凉、庆阳和咸阳等地市, 在陕西省高陵县注入渭河。泾河流域地貌复杂, 有山区、丘陵、高原、平原 4 种类型, 地形起伏较大, 如图 1 所示<sup>[6]</sup>。

统计学发现, 对数据进行分析时, 如果数据趋于正态分布, 那么利用数据所作的预测模型就相对精确。然而这些预测模型都是针对时间序列数据做的线性或非线性插值, 不涉及空间插值。降雨量数据经过变换后对于空间插值精度的影响以及变换方法不同是否精度也不同需要研究<sup>[7-8]</sup>。文献[9]在海河流域应用数据变换的方式提高了空间插值精度<sup>[9]</sup>。本文对泾河流域有效站点多年平均月降雨量分别进行平方根、立方根、对数变换, 然后根据上述泾河流域自然地理特征选择普通克里格法(Ordinary Kriging)和梯度距离平方反比法(GRDS)进行月降雨量插值。并和原数据插值结果进行比较。在比较空间插值精度时, 本文采用交叉验证(cross validation)的方法对各插值结果进行对比分析, 运用表征估计值可能的误差范围的平均绝对误差(MAE)、反映插值相对精确性的平均相对误差(MRE)、反映利用数据估值的灵敏度和极值效应的均方根误差(RMSE)三个参数评

表 1 四个测站 1951—2005 年月平均降水量 mm

测站	月份											
	1	2	3	4	5	6	7	8	9	10	11	12
53947	53	69	255	362	588	654	1191	1214	871	492	185	45
53929	69	88	234	374	539	725	973	1158	882	510	166	50
53923	45	68	196	337	530	731	1088	990	776	417	121	38
53915	33	53	138	281	459	714	1080	992	662	380	77	23

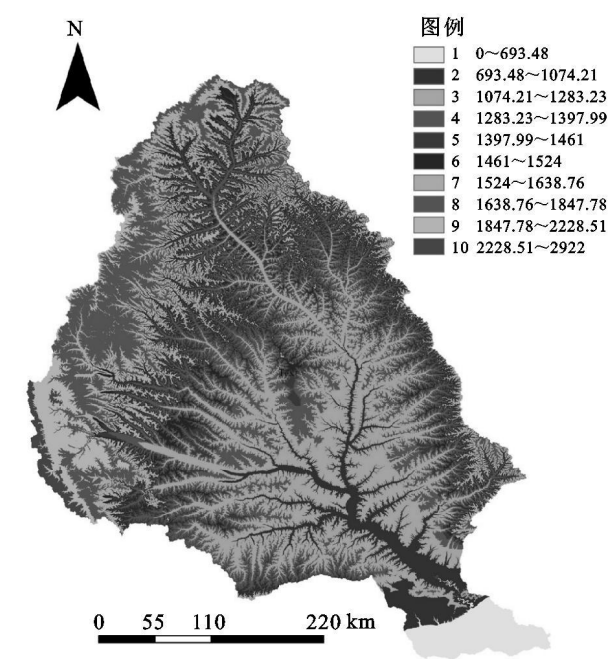


图 1 泾河流域 DEM 分级图

本文参阅相关文献采用立方根、平方根和对数 3 种变换

估插值方法效果。其计算方法如下式:<sup>[4-5,9]</sup>

$$MAE = \frac{1}{n} \sum_{i=1}^n |Z_i - Z'_i| \tag{1}$$

$$MRE = \frac{n}{\sum_{i=1}^n |Z_i|} \times MAE \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Z_i - Z'_i)^2}{n}} \tag{3}$$

式中:  $Z_i$  ——数据的真实值;  $Z'_i$  ——数据的估计值;  $n$  ——数据序列的个数。

首先说明数据来源并对数据做基本处理, 然后简要介绍两种插值方法的原理, 并对变换前后的插值结果进行比较。

2 数据说明及处理

本文所利用的泾河流域矢量边界图由泾河流域数据中心提供。降雨量数据从 1951—2005 年逐日气象要素数据库中抽取出来, 考虑到泾河流域内部的站点较少, 于是把陕甘宁 3 省的有效站点挑选出来, 利用 SQL Server 导出数据, 并用 Matlab 计算月均降雨量, 如表 1 所示, 限于篇幅这里只给出 4 个站点的月均降雨量。可以先利用这些站点插值, 然后用泾河流域矢量边界图截取插值结果图, 在一定程度上能够避免只利用泾河流域内部的几个站点插值造成的精度缺失。

对降雨数据进行预处理, 在数据分析的基础上对其进行数据变换方式的偏斜度优化选择, 然后研究处理前的数据和利用上述 3 种方法处理得到的数据在空间插值上的精度, 为选择相对最优的月均降雨量插值方法提供参考。偏斜度实际上是随机序列的 3 阶中心矩, 它的表达式如式(4)<sup>[10]</sup>。

$$S = \frac{E[X - E(X)]^3}{N} \tag{4}$$

式中:  $S$  ——偏斜度;  $E(X)$  —— $X$  数学期望;  $N$  ——数据个数。

偏斜度实际上表征了概率密度为非正态分布的随机序列偏离正态分布的歪斜程度, 其中正偏斜度表示不对称部分的分布更趋向正值, 负偏斜度表示不对称部分的分布更趋向负值。逐年月降雨量数据序列经过各种变换后求得的偏斜度, 如表 2 所示。

通过对比表 2 我们可以发现, 1, 2, 3, 4, 6, 12 月的数据经过对数变换后更趋向于正态分布; 7 月、8 月、9 月的数据没有变换前的要比任何变换的更趋向于正态分布; 5 月的数据经过平方根变换更趋向于正态分布; 10 月、11 月的数据经过立方根变换后的更趋向于正态分布。于是, 本文借鉴文献[9]提出的方法, 讨论每个月的降雨数据在插值方法不同、数据变换不同两种情况下的插值结果, 以获得精度相对较高的多年平均月降雨插值。

表 2 站点月降雨数据变换偏斜度分析

月 份	1	2	3	4	5	6	7	8	9	10	11	12
原数据	2. 87	2. 95	3. 98	0. 85	0. 61	0. 39	0. 13	− 0. 08	0. 55	0. 94	1. 02	2. 13
平方根变换	1. 55	1. 62	2. 13	− 1. 08	0. 44	− 0. 35	0. 44	0. 36	− 0. 60	1. 02	1. 55	1. 34
立方根变换	0. 91	1. 03	1. 33	0. 94	− 0. 89	0. 30	0. 37	0. 55	0. 69	0. 65	0. 89	1. 59
对数变换	0. 48	0. 55	0. 87	− 0. 37	0. 58	− 0. 22	0. 26	0. 33	0. 58	0. 88	1. 03	1. 13

3 空间插值方法及插值结果

空间插值属于地理统计学的范畴, 它根据获得的一定数量的反映了地理要素空间分布的全部或部分特征的空间样本, 对地理要素的未知空间特征进行预测, 即根据已知的空间数据估计未知空间的数据<sup>[3,11]</sup>。空间插值方法可以分为仅考虑平面位置影响的二维插值、仅考虑高程影响的降雨- 高程线性回归法以及同时考虑平面位置和高程影响的三维空间插值。二维插值方法有人工绘制降雨等值线法、泰森多边形法、距离平方反比加权平均法( IDSW) 和普通克里格法( Ordinary Kriging) 等, 三维空间插值方法中比较常见的是修正距离平方反比法( MRDS) 和梯度距离平方反比法( GRDS)<sup>[11]</sup>。

3.1 普通克里格法

克里格( Kriging) 法也叫做地理统计法, 实际上是一种普适的最小二乘法回归技术。其基础是变异函数理论和结构分析, 在有限区域内对区域化变量进行无偏最优估计。它把区域性未知变量看成是一组空间相关的随机变量。可以表达为

$$Z(\lambda) = \mu(\lambda) + \varepsilon(\lambda) + \xi(\lambda)$$
 (5)

式中:  $\mu(\lambda)$  —— 描述趋势性成分的函数, 也叫平均项;  $\varepsilon(\lambda)$  —— 反映空间变化结构的随机变量;  $\xi(\lambda)$  —— 剩余残差项。

实际上克里格法的基本思想类似于距离平方反比法, 即距离近的点之间的变量值应当比距离远的点之间的变量值更相似。不同的是, 克里格法采用半方差函数代替了距离平方反比法的距离函数来度量区域性变量的空间变异, 然后由观测值和待估值的半方差函数构建克里格线性方程组求解满足变量的估算方差最小且是无偏估计的权重系数。因此克里格法的核心是推求半方差函数, 其表达式为

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(\lambda_i) - Z(\lambda_i + h)]^2$$
 (6)

式中:  $N(h)$  —— 相距为  $h$  的数据对数;  $Z(\lambda_i)$  —— 位置  $\lambda_i$  的观测值。

半方差函数确定之后, 选用数学模型进行优化拟合, 推求最优的半方差函数。当趋势项  $\mu(\lambda)$  是未知常数时的克里格法叫做普通克里格法( OK: Ordinary Kriging)。ARCGIS 的空间分析模块提供了专门的克里格法分析工具<sup>[11]</sup>。

3.2 梯度距离平方反比法

梯度距离平方反比法( GRDS: Gradient plus reverse distance squared) 在距离权重的基础上考虑降雨随海拔和平面位置的变化。其表达式为

$$P^*(x, y, z) = \frac{\sum_{i=1}^N \frac{P(x_i, y_i, z_i) + C_y(y - y_i) + C_z(z - z_i)}{d_i^2}}{\sum_{i=1}^N \frac{1}{d_i^2}}$$
 (7)

式中:  $P^*(x, y, z)$  —— 待估点 $(x, y, z)$  处的雨量;  $N$  —— 参与计算的雨量站个数;  $P(x, y, z)$  —— 位于 $(x_i, y_i, z_i)$  处的第  $i$  个参与计算的雨量站的雨量;  $C_x$ 、 $C_y$  和  $C_z$  为降雨与平面位置  $x, y$  和海拔高程  $z$  的回归系数;  $d_i$  —— 待估点到第  $i$  个参与计算的雨量站点的距离。通过 ARCGIS 的二次开发工具 AO, 在 .NET 环境下编程实现梯度距离平方反比法以及交叉验证<sup>[11]</sup>。

3.3 结果分析

在对原数据做平方根、立方根、对数变换的基础上, 采用普通克里格法和梯度距离平方反比法进行多年平均降雨量的插值。得到的各月交叉验证结果如表 3 和表 4 所示。

由表 3 可知, 12 个月中仅有 5 月和 11 月采用普通克里格法对原数据进行降雨量插值效果较好, 其他月份都是对数变换优于原数据、平方根和立方根变换。表 4 中只有 4 月原数据优于各种变换, 其他月份对数变换的结果都优于原数据、平方根和立方根变换。于是可以得到这样的结论: 对泾河流域进行多年平均月降雨插值时, 将原数据进行对数变换并做空间插值时, 只有个别月份精度下降, 绝大多数月份精度都能得到提高。同时对比前述偏斜度分析结果可知, 虽然有些月份的原数据比对数变换更趋于正态分布, 但反映到空间插值结果上还是不如对数变换后的精确。比如 7 月, 8 月虽然原数据更趋于正态, 但由表 3 和表 4 可知两种插值方法都是在对数变换结果最优; 5 月的数据经过平方根变换更趋于正态, 10 月、11 月的数据经过立方根变换更趋于正态, 但这些月份两种插值方法的结果都是对数变换下最优。

由表 3 可以计算出普通克里格法的原数据、平方根变换、立方根变换和对数变换的平均绝对误差( MAE) 分别为: 5. 35, 5. 36, 5. 29, 5. 12; 平均相对误差( MRE) 分别为: 14. 17%, 13. 67%, 12. 33%, 9. 58%。由表 4 可以计算出梯度距离平方反比法的平均绝对误差( MAE) 分别为: 6. 52, 6. 41, 6. 26, 6. 07; 平均相对误差( MRE) 分别为 28%, 27%, 24. 25%, 22%。对比四组数据可以看出, 无论数据变换与否, 普通克里格法都要优于梯度距离平方反比法。以最优的对数变换为例, 普通克里格法的平均绝对误差( MAE) 为 5. 12, 平均相对误差为 9. 58%, 而梯度距离平方反比法的平均绝对误差为 6. 07, 平均相对误差为 22%。

由于篇幅所限, 这里给出泾河流域 6 月多年平均降雨量经过 3 种数据变换后的插值结果图与原数据的插值结果图( 见附图 1, 2)。附图 2 中原数据、平方根变换和立方根变换的插值结果有牛眼存在, 即使最优的对数变换插值结果图也不如附图 1 中原数据的插值结果图圆滑。可以直观的说明各种数据变换下普通克里格插值要优于梯度距离平方反比插值。

表 3 降雨量插值交叉验证(普通克里格法, 数据变换)

月 份	1	2	3	4	5	6	7	8	9	10	11	12
MAE 原数据	0.47	0.70	1.34	3.18	4.27	8.19	16.58	17.24	7.44	3.22	1.01	0.50
平方根变换	0.49	0.66	1.32	3.14	4.57	8.45	16.44	17.22	7.28	3.20	1.13	0.47
立方根变换	0.50	0.61	1.21	3.11	4.51	8.37	16.21	17.12	7.16	3.12	1.11	0.45
对数变换	0.44	0.55	1.18	2.97	4.38	8.15	15.46	16.88	6.99	2.97	1.07	0.37
MRE 原数据	0.22	0.24	0.18	0.14	0.11	0.12	0.12	0.15	0.14	0.11	0.04	0.13
平方根变换	0.24	0.21	0.17	0.13	0.10	0.11	0.11	0.12	0.12	0.10	0.11	0.12
立方根变换	0.23	0.21	0.16	0.13	0.09	0.09	0.09	0.09	0.12	0.09	0.09	0.09
对数变换	0.22	0.20	0.14	0.09	0.05	0.06	0.06	0.08	0.07	0.05	0.06	0.07
RMSE 原数据	0.52	0.74	1.56	2.88	5.55	15.2	21.32	24.6	5.67	3.17	0.79	0.67
平方根变换	0.56	0.75	1.55	2.67	5.60	15.4	21.22	23.5	5.49	3.14	1.21	0.64
立方根变换	0.54	0.39	1.49	2.34	5.34	14.9	21.31	22.5	5.27	3.04	1.12	0.51
对数变换	0.55	0.66	1.48	2.17	5.14	14.7	20.7	21.4	5.12	2.99	0.98	0.42

表 4 降雨量插值交叉验证(梯度距离平方反比法, 数据变换)

月 份	1	2	3	4	5	6	7	8	9	10	11	12
MAE 原数据	0.55	0.91	2.01	3.01	5.17	9.11	18.97	20.31	9.07	4.57	2.88	1.68
平方根变换	0.51	0.82	1.95	3.97	4.84	8.84	18.82	19.58	8.91	4.52	2.71	1.46
立方根变换	0.50	0.75	1.74	3.55	4.62	8.62	18.74	19.34	8.86	4.45	2.64	1.35
对数变换	0.45	0.61	1.27	3.34	4.41	8.41	18.55	19.22	8.54	4.26	2.55	1.21
MRE 原数据	0.37	0.29	0.24	0.11	0.19	0.27	0.25	0.31	0.33	0.29	0.34	0.37
平方根变换	0.34	0.27	0.21	0.22	0.17	0.26	0.24	0.29	0.31	0.27	0.32	0.34
立方根变换	0.32	0.24	0.18	0.18	0.14	0.22	0.22	0.24	0.29	0.24	0.31	0.33
对数变换	0.29	0.19	0.17	0.17	0.09	0.19	0.20	0.23	0.27	0.24	0.28	0.32
RMSE 原数据	0.78	1.24	2.14	3.01	7.24	18.3	28.44	30.62	7.24	4.06	2.01	1.04
平方根变换	0.72	1.21	1.94	3.32	7.12	17.9	27.69	28.47	7.20	4.01	1.96	1.01
立方根变换	0.68	0.92	1.81	3.21	6.97	15.8	26.77	27.61	7.11	3.97	1.84	0.94
对数变换	0.61	0.88	1.72	3.09	6.80	15.4	25.54	27.52	7.09	3.88	1.77	0.91

4 结 论

(1) 偏斜度分析虽然可以得出每个月经过何种变换更趋于正态分布,但这种选择对于空间插值意义不大。交叉验证结果表明虽然有些月份原数据或者平方根变换、立方根变换更能趋于正态分布,但反映到空间插值上还是对数变换后的结果精度较高。

(2) 虽然梯度距离平方反比法考虑了高程,但是在插值精度上无论何种变换都不如普通克里格法的高。交叉验证结果说明,利用对数变换选择普通克里格插值方法,可以有效地并能获得精度相对较高的多年平均月降雨量插值,为以后的日降水插值研究提供了较为可靠的本底文件,并能在一定程度上提高日降水插值的精度。

参考文献:

[ 1 ] 贾仰文,王浩,等. 分布式流域水文模型原理与实践 [ M ]. 北京: 科学出版社, 2005: 57-58.

[ 2 ] 何红艳,郭志华,肖文发. 降水空间插值技术的研究进展 [ J ]. 生态学杂志, 2005, 24( 10 ): 1187-1191.

[ 3 ] 朱求安,张万昌. 流域水文模型中面雨量的空间插值 [ J ]. 水土保持研究, 2005, 13( 4 ): 11-14.

[ 4 ] 赵登忠,张万昌,刘三超. 基于 DEM 的地理要素 PR

ISM 空间内插研究 [ J ]. 地理科学, 2004, 24( 2 ): 205-210.

[ 5 ] 朱求安. HEC 模型在汉江褒河流域的适应性研究 [ D ]. 南京: 南京大学研究生院, 2005.

[ 6 ] 王兮之,索安宁,洪军,等. 泾河典型流域水沙变化及其景观格局分析 [ J ]. 水土保持研究, 2006, 14( 8 ): 260-263.

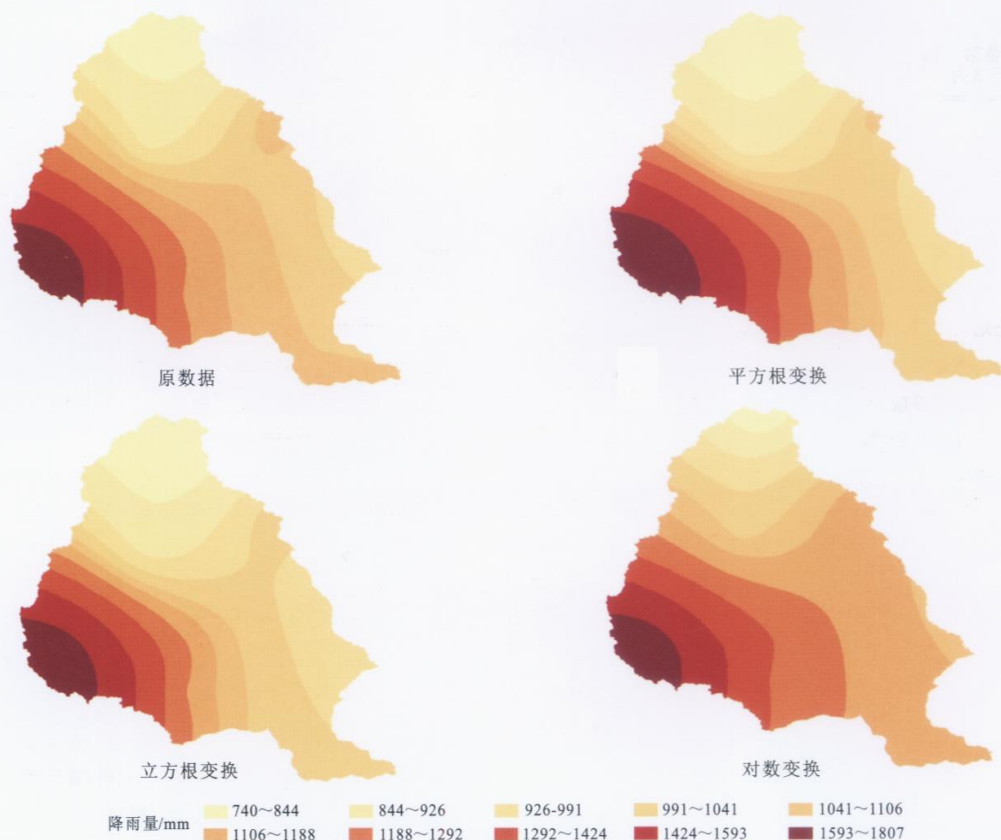
[ 7 ] Michael F Hutchinson. Interpolation of rainfall data with thin plate smoothing splines - Part I: tow dimensional smoothing of data with short range correlation [ J ]. Journal of Geographic Information and Decision Analysis, 1998, 2 ( 2 ): 139-151.

[ 8 ] Genton M G. Analysis of rainfall data by robust spatial statistic using S+ SPATIALSTATS [ J ]. Journal of Geographic Informationand Decision Analysis, 1998, 2 ( 2 ): 116-126.

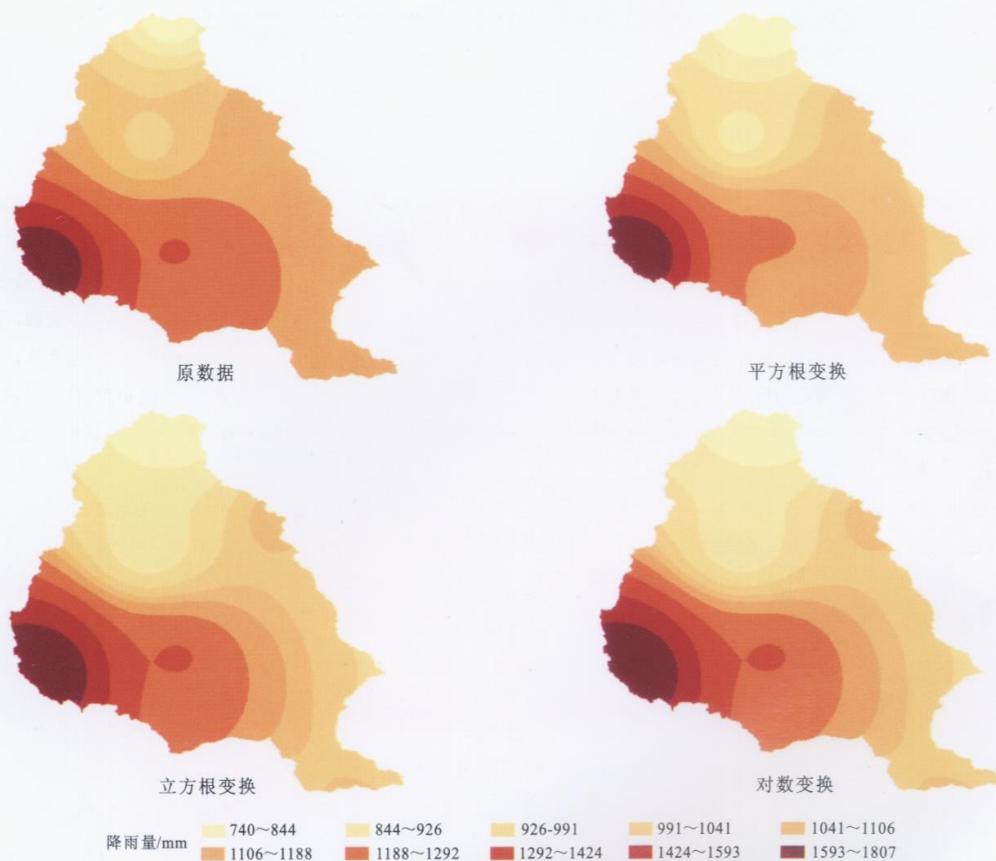
[ 9 ] 刘登伟,封志明,杨艳昭,等. 海河流域降水空间插值方法的选取 [ J ]. 地球信息科学, 2006( 12 ): 75-79.

[ 10 ] 谢忠秋,丁兴烁. 应用统计学 [ M ]. 上海: 立信会计出版社, 2005: 53-57.

[ 11 ] 黄杏元,马劲松,汤勤. 地理信息系统概论 [ M ]. 北京: 高等教育出版社, 2005: 93-97.



附图1 普通克里格法插值结果图（径河，6月）



附图2 GRDS插值结果图（径河，6月）