

文章编号: 1008-2786-(2016)5-591-08

DOI: 10.16089/j.cnki.1008-2786.000166

# 基于 PLS 回归模型的石羊河流域降水空间分布

董慧慧<sup>1,2</sup>, 上官冬辉<sup>1\*</sup>

(1. 中国科学院寒区旱区环境与工程研究所/冰冻圈科学国家重点实验室, 甘肃 兰州 730000;

2. 中国科学院大学 研究生院, 北京 100049)

**摘 要:** 利用石羊河流域内国家气象站点和中国内陆河湖水文站点 2010—2012 年逐日降水观测资料, 基于 DEM 资料提取流域经度、纬度、坡度、坡向、海拔等地形因子, 建立降水量与地理地形因子之间的多元偏最小二乘 (partial least squares, PLS) 回归模型, 对石羊河流域 2010—2012 年在年、季、月尺度上的降水分布特征进行研究, 结果显示: 1) 采用交叉验证法对 PLS 与反距离加权法 (IDW)、样条函数法 (SPLINE)、普通克里金插值法 (OK) 等传统插值方法的精度验证结果比较发现, PLS 的平均绝对误差 (MAE) 和均方根误差 (RMSE) 值均是最小的。PLS 回归模型对降水量的拟合相关系数除 2011/2012 年冬季分别为 0.68 和 0.66, 其余模型的拟合相关系数均在 0.81 以上, 且显著性水平在 95% 以上。2) 不同的时间及时空尺度, 地形因子对降水量的影响是不同的, 基本上海拔和坡度对降水量的影响是较大的, 且随海拔和坡度的增加降水量增加。3) 石羊河流域 2010—2012 年年降水量从南向北逐渐递减, 最大降水量均在南部祁连山区; 季节降水量大小依次为夏季 > 秋季 > 春季 > 冬季, 夏季降水量较丰富; 不同年的夏季 6/7/8 月的降水特征也各不相同, 2011 年 7 月和 2012 年 8 月降水量相对较丰富。

**关键词:** 石羊河流域; DEM; PLS; 降水量

**中图分类号:** P343, P426.6

**文献标志码:** A

石羊河是甘肃省河西走廊三大内陆河流域之一, 由此形成的绿洲, 是防止北部巴丹吉林沙漠和腾格里两大沙漠汇合, 保卫河西走廊及我国西北部的重要生态区。近些年由于自然演变以及人类活动的叠加影响, 石羊河流域源头的祁连山冰川逐渐萎缩、雪线上升, 上游来水量逐年减少, 下游民勤盆地的地下水水位下降, 水质矿化, 土地沙漠化严重, 流域面临严重的水资源问题<sup>[1]</sup>。

因此针对石羊河流域的降水量时空分布特征, 研究其对该区域水资源管理及水资源合理配置, 具有十分重要的指导意义。

## 1 研究现状

目前, 获取降水量分布的方法主要有基于雨量站点进行降水量插值和卫星降水数据产品。基于站点的插值方法包括: 反距离加权法 (IDW)、样条函数法 (SPLINE)、普通克里金插值法 (Ordinary Kriging) 及克里金系列插值法等经典插值方法, 以及近些年逐渐受到重视的多元统计回归方法, 如 Daly 的 PRISM 插值算法<sup>[2]</sup>、舒守娟的偏最小二乘插值算法<sup>[3]</sup>、周锁铨的逐步插值算法<sup>[4]</sup>等。卫星降水数据产品国际上已发布的有 CMORPH 降水产品<sup>[5]</sup>、

收稿日期 (Received date): 2015-07-23; 改回日期 (Accepted): 2015-09-02。

基金项目 (Foundation item): 科技部重大科学研究计划 (2013CBA01808), 国家自然科学基金项目 (41271082, 41130638)。[Major Research Program from the Ministry of Science and Technology (2013CBA01808); National Nature Science Foundation of China (41271082; 41130638).]

作者简介 (Biography): 董慧慧 (1991-), 女, 河南周口人, 在读硕士研究生, 主要从事山区降水研究。[Dong Huihui, female, born in Zhoukou, Henan, M.Sc. candidate, interested in precipitation research in mountain area.] E-mail: donghui@lzb.ac.cn

\* 通信作者 (Corresponding author): 上官冬辉 (1975-), 男, 副研究员, 主要从事冰川遥感监测方面研究。[Shangguan Donghui (1975-), male, associate professor, mainly engaged in the glacier monitor with remote sensing.] E-mail: dhguan@lzb.ac.cn

TRMM 降水资料<sup>[6]</sup>和 APHRODITE 数据产品<sup>[7]</sup>等比较成熟的产品,国内的有 FY2C 卫星降水产品<sup>[8]</sup>等。在石羊河流域,陈贺采用基于最小二乘法的多元统计回归法对流域上游的一次降水过程的日降水数据进行插值分析<sup>[9]</sup>,马金辉利用 TRMM 降水资料进行降尺度处理获得了上游 10 a 平均降水的分布结果<sup>[10]</sup>。事实上,日降水尺度或者多年平均的降水特性研究对分析年内降水空间分布缺乏认识,且由于石羊河流域雨量站点大多集中在上游地区,下游站点稀少,无论是站点插值还是卫星降水产品对石羊河整个流域在年、月尺度上较为精确的降水分布研究都很少。本文尝试用多元统计回归插值法对石羊河整个流域 2010—2012 年的降水量空间分布进行研究分析。

多元统计回归插值法是建立气象要素与地理地形因子之间的回归方程,在进行待估点的插值计算时依赖的是该位置上各因子对气象要素的影响大小,对站点的密集程度要求不高。该方法因考虑了气象要素的空间分布与地理地形要素之间的关系,在定量反映真实地形各要素的空间分布上有很大优势<sup>[3]</sup>。多元回归分析方法包括有基于普通最小二乘算法的回归分析、主成分回归分析、逐步回归分析、偏最小二乘回归分析等。但当多个自变量之间存在多重相关性时,采用普通最小二乘算法进行建模时很难建立稳健可靠的模型;逐步回归法在迭代剔除对因变量影响不显著的变量过程中,难以客观选择合适的变量参与建模并对因变量进行合理解释;偏最小二乘回归分析法则吸收了主成分回归提取成分的思想,先从自变量集合中提取成分,所提取的成分是线性无关的,解决了模型中自变量的多重相关问题,同时,对自变量与因变量进行相关分析,既考虑到成分对自变量的概括性,也实现了对因变量的解释意义<sup>[11]</sup>。舒守娟等采用偏最小二乘回归统计方法,利用气象观测站点 45 a 平均降水资料,建立降水分布与地理地形因子的关系,研究了中国区域降水分布特点<sup>[3]</sup>;郭忆则采用 2001—2010 年 10 a 年降水资料,利用偏最小二乘法对长江中下游平原地区的降水量空间分布进行估算与分析<sup>[12]</sup>。但对于该方法在小流域及更小时间尺度上的降水插值应用上还没有出现过。

综合对石羊河流域雨量站点的分布特征及地形因子对降水量影响等多方面因素的考虑,本文选择偏最小二乘回归方法对石羊河流域的降水量分布进

行插值研究尝试。以石羊河流域内 2010—2012 年雨量站点的逐日降水资料为基础,基于 DEM 提取经度、纬度、坡度、坡向、高程等地理地形因子的栅格图层,建立降水量与各因子之间的多元 PLS 回归模型,并将插值结果与传统插值方法(IDW, SPLINE, OK)进行精度比较,筛选出最适合石羊河流域降水分布研究的方法,并分析石羊河流域的年、季、月尺度降水分布特征,为后续研究该区域水文水资源及旱涝分析提供基础。

## 2 方法和数据

### 2.1 研究区概况

石羊河流域位于甘肃省河西走廊东部,地理位置为  $101^{\circ}07' \sim 104^{\circ}16'E$ ,  $37^{\circ}07' \sim 39^{\circ}27'N$ ,流域面积为  $4.16 \times 10^4 \text{ km}^2$ ,东与腾格里沙漠接壤,南部靠祁连山北麓,北于巴丹吉林沙漠相接,西止大黄山,地势南高北低,自西南向东北倾斜,海拔  $1\,247 \sim 4\,960 \text{ m}$ 。受流域南部山区山脉阻挡,石羊河流域有明显的景观分带现象,自南向北呈现出上游祁连山区、中游河西走廊平原区、下游低山丘陵和荒漠区。流域水系发源于祁连山,自东向西由大靖河、古浪河、黄羊河、杂木河、金塔河、西营河、东大河、西大河等 8 条河流组成,河流补给源主要为山区大气降水和高山冰雪融水<sup>[13]</sup>(图 1)。

### 2.2 资料来源

本文所用降水数据来源于中国气象科学数据共享服务网(<http://cdc.cma.gov.cn/>)及中国水文年鉴的甘肃省内陆河湖水文资料,数据内容为石羊河流域内 4 个国家气象站点和 26 个水文站 2010—2012 年逐日降水量资料,站点分布见图 1。区域南部山区地势起伏较大,站点分布较多,主要集中在海拔  $2\,000 \sim 3\,000 \text{ m}$ ;北部地势起伏较小,站点也较少。本文采用的 DEM 数据是 STRM DEM V4 数据,空间分辨率为  $90 \text{ m}$ ,数据来源于 CGIAR - CSI(<http://srtm.csi.cgiar.org>)。在数据预处理时,将 DEM 重采样为  $1 \text{ km}$  分辨率,并在 ARCGIS 软件平台上提取同等分辨率的经度、纬度、坡度、坡向等要素的栅格图层。

## 3 研究方法

### 3.1 基于 PLS 的多元线性回归建模

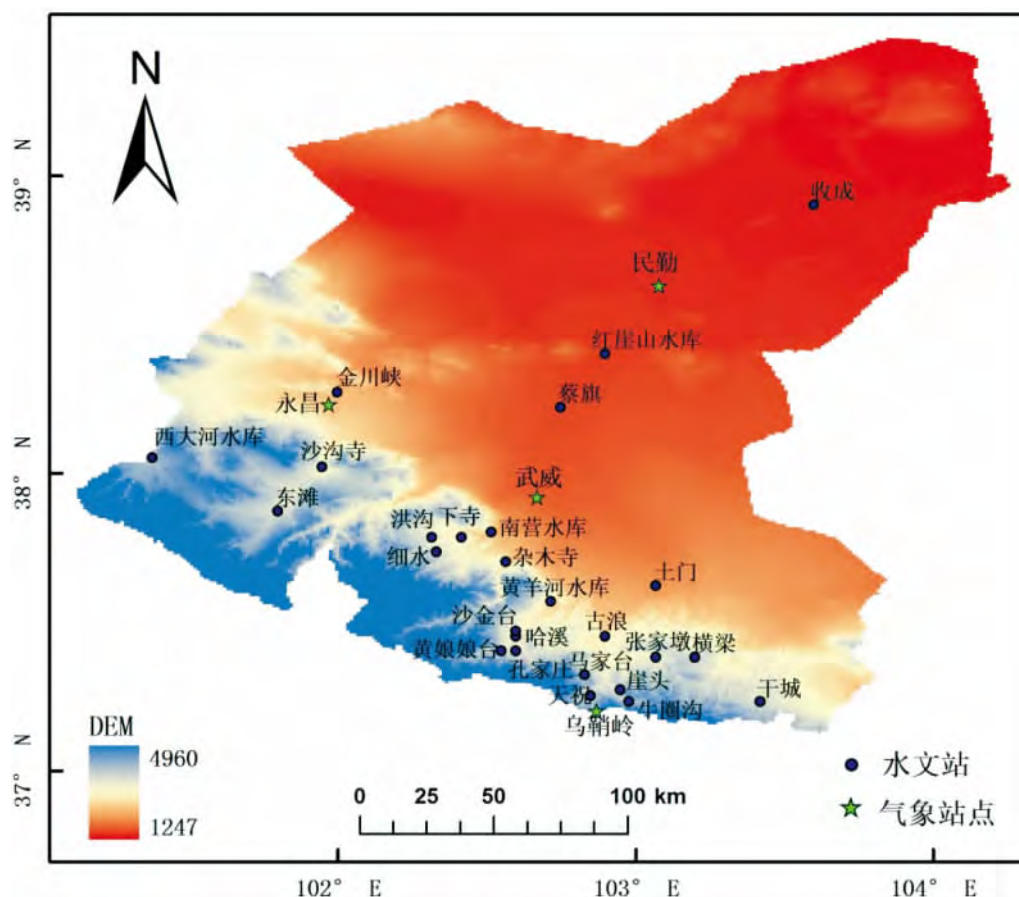


图1 石羊河流域 DEM 及站点分布

Fig. 1 DEM of Shiyang River Basin and distribution of gauge stations

地形对局地气候的影响,在一定程度上可与大气环流、太阳辐射的作用相比拟,石羊河流域降水很大程度上受地形影响<sup>[9]</sup>。因此降水量可描述为:降水量  $P = P(\text{lat}, \text{lon}, h, \text{aspect}, \text{slope}, \varepsilon)$ , 其中  $\text{lat}$ 、 $\text{lon}$  分别为纬度、经度;  $h$  为高程;  $\text{aspect}$  为地形坡向;  $\text{slope}$  为地形坡度;  $\varepsilon$  为回归分析的残差。降水量与各因子之间的多元线性回归模型方程表达式为

$$P = a_0 + a_1 * \text{lat} + a_2 * \text{lon} + a_3 * \text{slope} + a_4 * \text{aspect} + a_5 * h + \varepsilon \quad (1)$$

式中  $a_0$  为常数项,  $a_1 \sim a_5$  为对应各项地理地形因子的系数。

偏最小二乘回归法 (partial least squares, PLS) 是由伍德 (S. Wold) 和阿巴诺 (C. Albano) 等人首次于 1983 年提出的一种新型的多元统计分析方法<sup>[14]</sup>。近几十年来,它在理论、方法和应用方面都得到了迅速的发展。该方法的优势在于能够在一个算法下同时实现多元线性回归建模、自变量数据结构简化以及两组变量之间的相关性分析等。具体的算法思想许多文章和书籍中都有提到<sup>[3, 10]</sup>, 本文不

再详述。

### 3.2 成分提取原则

PLS 法通过提取成分的方式确定模型,但并不是提取的成分越多越好,当达到最佳分量数后,增加新的成分会降低模型的预测能力。最佳成分数可以通过交叉验证法来确定,交叉验证即每个观测值既参与模型的建立,又参与模型的评价,以此求得观测值的预测误差平方和 (Prediction Residual Error Sum of Squares, PRESS)

$$PRESS_m = \sum_{i=1}^n (y_i - \hat{y}_{m(-i)})^2 \quad (2)$$

其中  $n$  为样本数,  $m$  为提取的成分个数,  $y_i$  为因变量原始数据,  $\hat{y}_{m(-i)}$  是在建模时剔除样本点  $i$ , 取  $m$  个成分回归建模后,再用此模型计算  $y_i$  的拟合值。PRESS 体现了观测点的变动所引起的扰动误差和拟合方程的预测误差,其值越大说明所建模型越不稳定。在 PLS 回归分析过程中按照预测误差平方和最小的原则确定成分个数。以 2010—2012 年夏季的降水量为例,图 2 中列出了提取不同的分量计算出的 PRESS 值,并以图表的形式直观展示了 PLS 模

型应提取的最佳成分数: 2010—2012 年夏季分别提取 5 个、2 个、4 个成分时预测误差平方和最小, 此时模型的预测能力是最大的。

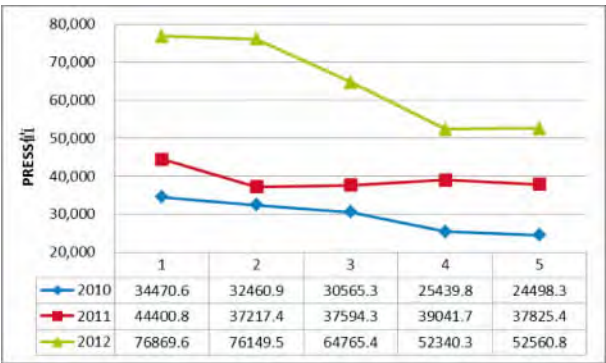


图2 基于 PLS 提取不同成分数的 2010—2012 年夏季降水量预测误差平方和

Fig. 2 The PRESS of summer precipitation (2010—2012) by extracting different fractions based on PLS model

3.3 降水量 PLS 回归模型

依据 PRESS 最小原则提取分量,生成降水量的 PLS 回归模型,表 1 列出了各 PLS 回归模型系数,其中  $a_0$  是常数项,  $a_1 \sim a_5$  分别对应纬度、经度、坡度、坡向、高程项的系数。 $R^2$  表示模型预测值和实测值之间的相关系数平方,体现了模型对数据的拟合优度,其值越大说明模型对数据的拟合越好。从表中可以看出,除 2011/2012 年冬季模型的拟合相关系数分别为 0.68(  $\sqrt{0.47}$  )、0.66(  $\sqrt{0.44}$  ),模型的拟合相关系数均在 0.81(  $\sqrt{0.66} = 0.81$  ) 以上,其中 2010 年年降水量模型拟合相关系数平方  $R^2$  达 84%, 2011 年和 2012 年年降水量模型对数据的拟合相关系数平方则达到了 90%, 且回归模型的显著性水平在 95% 以上,表明多元 PLS 回归算法对该区域降水模型建立具有一定的适用性。图 3 为回归模型中各因子的标准化回归系数直方图,系数的正负反映因子与降水量之间的正负相关关系,大小反映对降水量的影响大小,通过直方图可以直观看出每一个自变量因子在解释因变量时的重要性及相关程度。从直方图来看,海拔对降水量的影响是最大的,且海拔与降水呈正向相关关系,随海拔升高降水增加,这也造成在石羊河流域南部祁连山区降水量较为丰富,成为石羊河流域的主要水源地;其次,坡度对降水量的影响也比较显著,降水量与坡度呈正比,地形起伏变化越大的地方降水量较多,高程和坡度的增大有利于降水的形成。除此之外,2011 年降水

量受纬度、坡向因子的影响也较为显著,且均为负相关因子,随纬度增加,降水量减少,与石羊河流域从南到北的地貌形态一致。

表 1 石羊河流域 2010—2012 年降水量的 PLS 模型回归方程系数及  $R^2$

Tab. 1 Coefficients and  $R^2$  of PLS regression models for precipitation estimation in Shiyang River Basin in 2010—2012

时间	$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$R^2$
2010 年	-3 808.04	9.52	31.35	15.06	-0.16	0.23	0.84
2010 春	185.8	-36.58	11.73	5.15	-0.063	0.036	0.72
2010 夏	-130.37	17.37	-6.7	4.85	-0.03	0.117	0.85
2010 秋	-2 160.82	5.07	18.91	3.74	-0.05	0.05	0.73
2010 冬	49	-2.21	0.4	1.29	-0.017	0.0016	0.69
2010 6 月	-746.1	-1.85	7.33	1.96	-0.012	0.046	0.79
2010 7 月	-72.54	9.3	-3.25	3.08	-0.013	0.035	0.8
2010 8 月	688.27	9.91	-10.78	-0.18	-0.002	0.035	0.86
2011 年	461.51	-70.77	21.1	25.91	-0.21	0.18	0.9
2011 春	616.66	-12.82	-1.56	2.06	-0.026	0.041	0.81
2011 夏	2147	-56.3	0.38	14.16	-0.12	0.05	0.84
2011 秋	848.73	-47.22	9.21	7.07	-0.035	0.04	0.84
2011 冬	-2 673	-3.47	1.59	1.05	-0.011	0.001	0.47
2011 6 月	-1 960.52	7.26	15.83	2.35	-0.0001	0.05	0.8
2011 7 月	1 152.66	-11.05	-7.02	3.93	-0.03	0.01	0.74
2011 8 月	1 061.68	-28.53	0.71	7.88	-0.09	0.02	0.68
2012 年	-9 374.88	-19.18	95.37	15.28	-0.09	0.29	0.9
2012 春	681.11	-27.75	3.48	4.54	0.028	0.026	0.66
2012 夏	-6 120.27	5.7	55.46	5.66	0.0002	0.18	0.87
2012 秋	-1 327.7	-26.31	22.32	2.85	-0.04	0.05	0.88
2012 冬	347.76	-2.6	-2.37	0.68	-0.014	0.002	0.44
2012 6 月	-805.76	-3.76	8.82	0.2	-0.018	0.044	0.68
2012 7 月	-4 072.83	13.27	33.73	1.91	0.01	0.09	0.81
2012 8 月	920.28	-25.17	0.17	4.06	0.03	0.024	0.66

4 结果与分析

4.1 模型精度验证方法

为了对模型的降水误差进行分析,进一步说明 PLS 方法对石羊河流域降水空间插值的效果,以 2010—2012 年年降水数据为例,选择了 ARCGIS 软件中的几种插值方法如 IDW、OK、Spline 对数据插值,并用交叉验证法对 4 种插值方法的精度进行比较。

以平均绝对误差 (MAE) 和均方误差平方根 (RMSE) 作为交叉验证的检验指标,MAE 可以评估



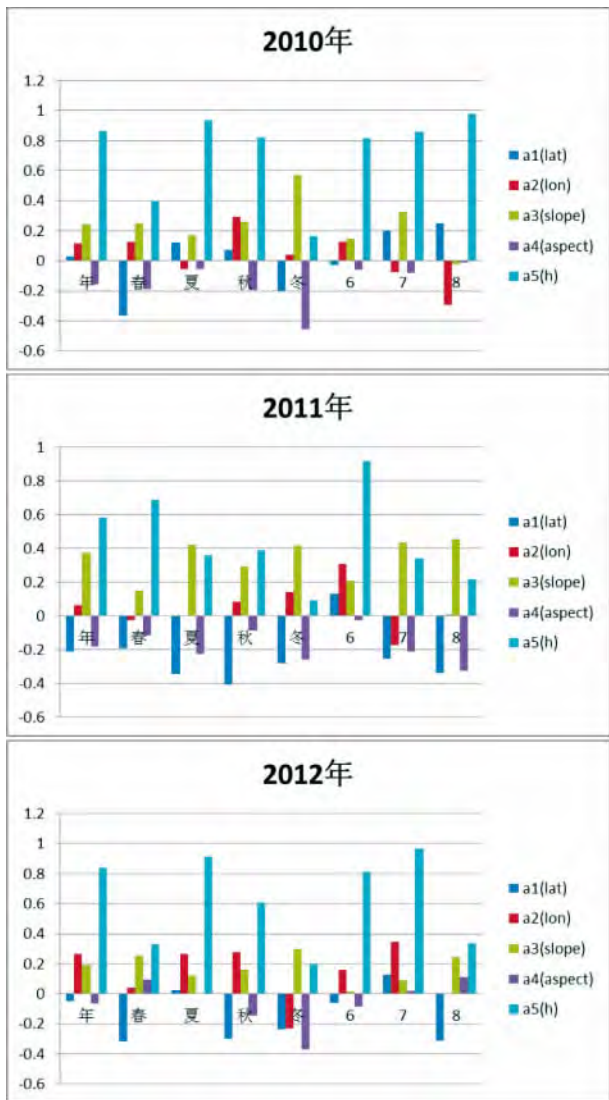


图3 2010—2012 年降水影响因子的标准化回归系数直方图  
Fig.3 Normalized regression coefficient of factors  
for precipitation in Shiyang River Basin in 2010—2012

估算值可能的误差范围 , $RMSE$  可以反映利用样点的估算灵敏度。 $MAE$ 、 $RMSE$  的值越小 ,表明模型模拟的精度越高。表达式如下:

$$MAE = \frac{\sum_{i=1}^n |P_{oi} - P_{ei}|}{n} \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_{oi} - P_{ei})^2}{n}} \tag{4}$$

其中  $P_{oi}$  为第  $i$  个站点的降水量观测值  $P_{ei}$  为第  $i$  个站点的降水量预测值  $n$  为参与检验的站点数。

4.2 误差分析

采用交叉验证法对 4 种插值方法的精度进行比较 ,结果如表 2 所示。从表中看出 ,PLS 模型的  $MAE$  和  $RMSE$  值是四种方法中最小的 ,OK 模型以

及 SPLINE 模型次之 ,IDW 法插值结果  $MAE$  和  $RMSE$  值是四种方法中最大的 ,插值效果优劣依次是  $PLS > OK > SPLINE > IDW$ 。这说明在石羊河流域降水观测站点分布不均匀 ,流域地形呈现南高北低 ,地形较为复杂的情况下 ,仅依赖站点之间距离权重获得的降水插值结果不及考虑了地理地形因子的多元 PLS 回归模型。

表 2 检验站点 2010—2012 年的交叉验证结果

Tab.2 Results of cross-validation error in 2010—2012 by four interpolation methods /mm

年	PLS		IDW		OK		SPLINE	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
2010	53.2	64.3	70.5	87.0	54.6	74.9	64.3	80.4
2011	45.9	58	74.2	92.7	55.4	74.6	64.5	82.2
2012	49.9	65.7	72.9	93.6	58.3	75.8	68.4	85.9

4.3 降水分布特征

利用 PLS 回归分析的方法建立降水量与地理地形因子之间的多元回归模型 ,并生成石羊河流域 2010—2012 年年、季、夏季 6/7/8 月的降水量分布图(图 5)。从图中可以看出 ,无论是年、季降水量 ,还是月降水量 ,降水量基本与海拔的走势一致 ,最大降水量均在南部祁连山区高海拔处 ,年最大降水量均在 850 mm 左右 ,2010 年降水量略小 ,2011 年、2012 年降水量相对丰富些 ,年降水量在南部祁连山区变化较大 ,但在中部平原区向北至巴丹吉林沙漠降水量变化不大且降水量较少。降水季节性差异较大 ,降水量大小依次为夏季 > 秋季 > 春季 > 冬季 ,夏季受西南暖湿气流影响 ,石羊河流域处于有利的降水环境中 ,夏季降水量较为丰富 ,占全年降水量的 50% 左右;而春季降水主要受青藏高原东部季风影响 ,秋季降水则受暖湿气流的微弱影响 ,但前者影响要远远弱于后者 ,因此秋季降水要比春季降水多;冬季受西伯利亚大陆冷高压影响 ,气流干燥寒冷 ,南部暖湿气流经过长途跋涉已经势力衰弱 ,又受祁连山脉阻挡 ,使得冬季降水稀少。而夏季三月份 6/7/8 月的降水量特征 2010 年夏季三个月的降水量相对 2011/2012 年总体较少 2011 年 8 月及 2012 年 7 月降水量较丰富。

5 结 论

本文利用石羊河流域 2010—2012 年内陆河水文

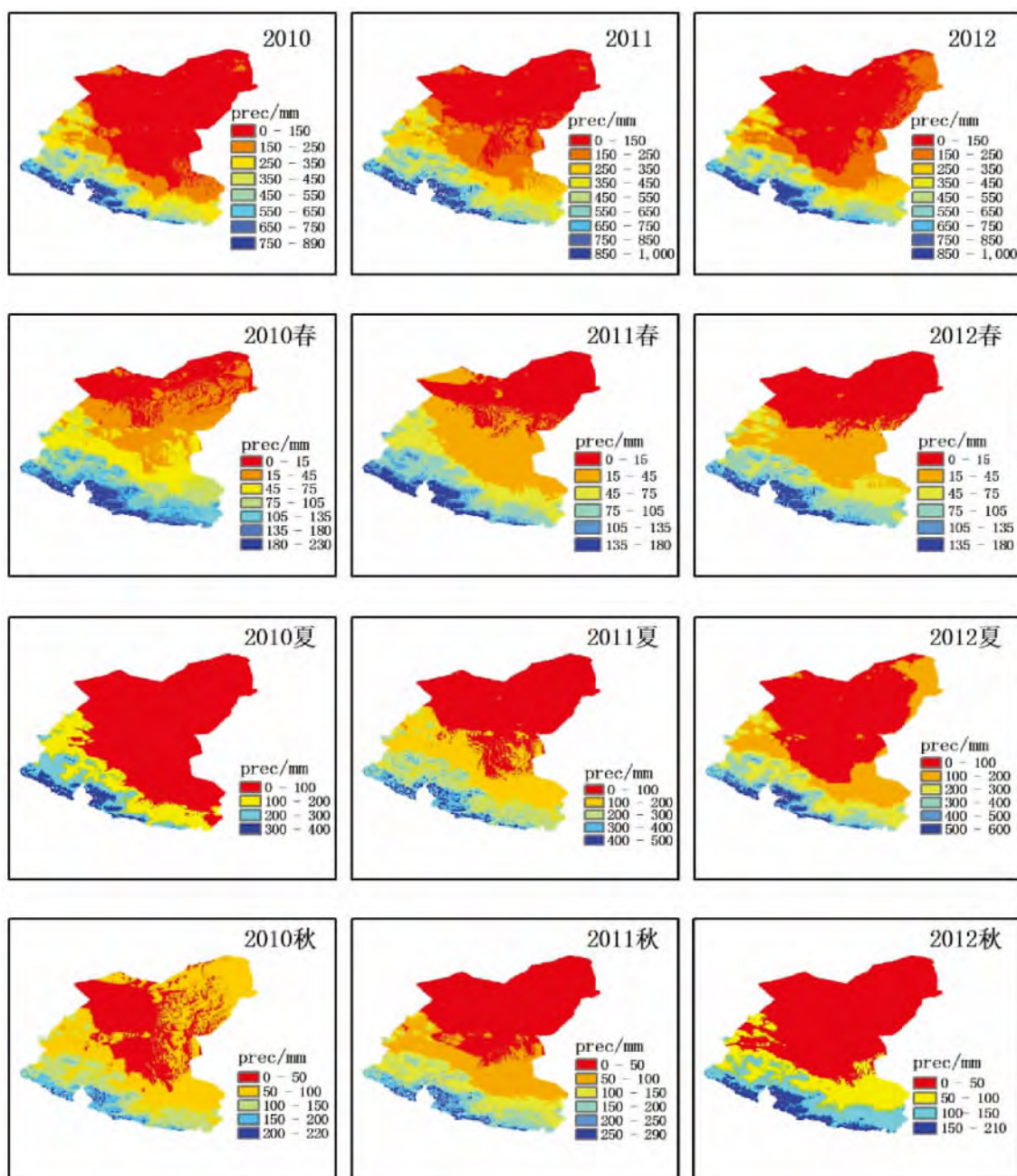
站点数据和国家气象站点数据,结合研究区 DEM 资料,选择经度、纬度、坡度、坡向、海拔等因子进行多元线性回归建模,采用可以解决变量间线性相关性,不剔除任何解释变量,对结果可以定性解释的偏最小二乘法对石羊河流域降水分布特征进行插值研究,结论如下:

1) 采用交叉验证的方法,通过与 IDW、SPLINE、OK 等传统插值方法的误差结果比较发现,由于降水观测站点分布不均匀(南多北少),且南部山区地形复杂,PLS 回归算法因考虑了各种地理地形因素对降水的影响,以及多元统计回归法对站点分布依赖性不高的优势,使得插值结果精度更高,

MAE 和 RMSE 是四种方法中最小的。

2) 不同的时间及时间尺度,地形因子对降水量的影响是不同的,基本上海拔和坡度对降水量的影响是较大的,且随海拔和坡度增加降水量增加,此外,2011 年降水量影响因子中,纬度和坡向对降水量影响也较大,且与降水量呈负相关关系。

3) 基于 PLS 回归模型的石羊河流域 2010—2012 年降水量模拟结果显示:石羊河流域年降水量呈现从南向北依次减少趋势,最大降水量分布在南部祁连山区高海拔地带,季节降水量大小依次为夏季 > 秋季 > 春季 > 冬季,2011 年 8 月及 2012 年 7 月降水量相对其他月较丰富。





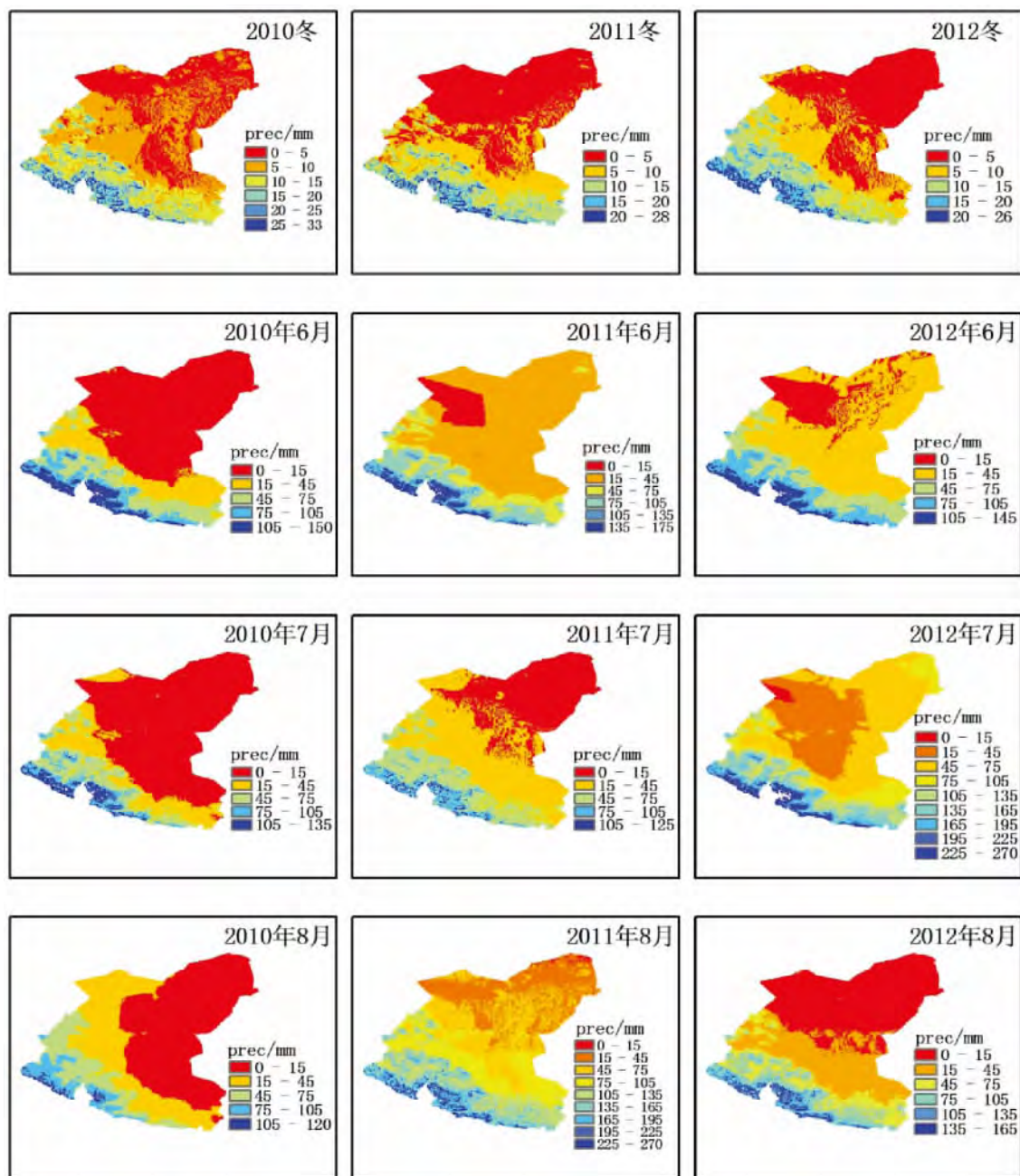


图4 石羊河流域2010—2012年的年、季、月降水量分布图

Fig. 4 Precipitation distribution of Shiyang River Basin in annual, season and month scale (2010—2012)

## 参考文献(References)

- [1] Ma Guojun, LIU Jundi, Lin Dong, et al. Status of water use and its econ-environmental effects in Shiyang River Basin[J]. Journal of Desert Research 2008 28(3): 592–597 [马国军, 刘君娣, 林栋, 等. 石羊河流域水资源利用现状及生态环境效应[J]. 中国沙漠 2008 28(3): 592–597]
- [2] Daly C, Neilson R P, Phillips D L. A statistical-topographic model for mapping climatological precipitation over mountainous terrain [J]. Journal of Applied Meteorology, 1994 (33): 140–158
- [3] Shu Shoujuan, Wang Yuan, Xiong Anyuan. Estimation and analysis for geographic and orographic influences on precipitation distribution

- in China [J]. Chinese Journal of Geophysics, 2007 50(6): 1703–1712 [舒守娟, 王元, 熊安元. 中国区域地理、地形因子对降水分布影响的估算和分析[J]. 地球物理学报, 2007 50(6): 1703–1712]
- [4] Zhou Suoquan, Xue Genyuan, Zhou Lifeng, et al. The stepwise interpolation approach of precipitation for spatial analysis based on GIS [J]. Acta Meteorologica Sinica 2006 64(1): 100–111 [周锁钎, 薛根元, 周丽峰, 等. 基于 GIS 降水空间分析的逐步插值方法[J]. 气象学报 2006 64(1): 100–111]
- [5] Xu Shiguang, Niu Zheng, Shen Yan et al. A research into the characters of CMORPH remote sensing precipitation error in China [J].

- Remote Sensing Technology and Application, 2014, 29(2): 189 – 194 [许时光, 牛铮, 沈艳, 等. CMORPH 卫星降水数据在中国区域的误差特征研究[J]. 遥感技术与应用, 2014, 29(2): 189 – 194]
- [6] He Huizhong, Cui Zhehu, Cheng Minghu, et al. TRMM Satellite and application of its products[J]. Meteorological Science and Technology, 2004, 32(1): 13 – 18 [何会中, 崔哲虎, 程明虎, 等. TRMM 卫星及其数据产品应用[J]. 气象科技, 2004, 32(1): 13 – 18]
- [7] Han Zhenyu, Zhou Tianjun. Assessing the quality of APHRODITE high-resolution daily precipitation dataset over contiguous China[J]. Chinese Journal of Atmospheric Sciences, 2012, 36(2): 361 – 373 [韩振宇, 周天军. APHRODITE 高分辨率逐日降水资料在中国大陆地区的实用性[J]. 大气科学, 2012, 36(2): 361 – 373]
- [8] Shi Chunxiang, Xie Zhenghui. A time downscaling scheme of precipitation by Using Geostationary Meteorological Satellite data[J]. Progress in Geography, 2008, 27(4): 15 – 22 [师春香, 谢正辉. 基于静止气象卫星观测的降水时间降尺度研究[J]. 地理科学进展, 2008, 27(4): 15 – 22]
- [9] Chen He, Li Yuanyuan, Yang Zhifeng, et al. Research on the relationship between terrain factors and precipitation[J]. Research of Soil and Water Conservation, 2007, 14(1): 119 – 122 [陈贺, 李原园, 杨志峰, 等. 地形因素对降水分布影响的研究[J]. 水土保持研究, 2007, 14(1): 119 – 122]
- [10] Ma Jinhui, Qu Chuang, Zhang Haixiao, et al. Spatial downscaling of TRMM precipitation data based on DEM in the upstream of Shiyanghe River Basin during 2001—2010[J]. Progress in Geography, 2013, 32(9): 1423 – 1432 [马金辉, 屈创, 张海筱, 等. 2001—2010 年石羊河流域上游 TRMM 降水资料的降尺度研究[J]. 地理科学进展, 2013, 32(9): 1423 – 1432]
- [11] Wang Huiwen. Partial Least – Squares regression model and applications[M]. Beijing: National Defence Industry Press, 1999. [王惠文. 偏最小二乘回归方法及其应用[M]. 北京: 国防工业出版社, 1999.]
- [12] Guo Yi, Bi Shuoben, Zhang Yuqing, et al. Estimation and analysis for geographic and orographic influences on precipitation distribution on the Yangtze Plain, middle and lower[J]. Science Technology and Engineering, 2013, 13(29): 8555 – 8560 [郭忆, 毕硕本, 张余庆, 等. 地理地形因子对长江中下游平原降水空间分布影响的估算和分析[J]. 科学技术与工程, 2013, 13(29): 8555 – 8560]
- [13] Xu Zongxue, Li Zhanling, Shi Xiaokun. Long-term trends of major climatic variables and runoff in the Shiyang River Basin[J]. Resources Science, 2007, 29(5): 121 – 128 [徐宗学, 李占玲, 史晓鲲. 石羊河流域主要气象要素及径流变化趋势分析[J]. 资源科学, 2007, 29(5): 121 – 128]
- [14] Word S, Ruhe A, et al. The collinearity problem in linear regression, the partial least square(PLS) approach to generalized inverses[J]. J stat Comp, 1984, 5: 735 – 743

## Spatial Distribution of Precipitation in Shiyang River Basin Based on PLS Regression Model

DONG Huihui<sup>1, 2</sup>, SHANGGUAN Donghui<sup>1</sup>

(1. State Key Laboratory of Cryospheric, cold and Arid Regions environmental and Engineering Research Institute, Chinese Academy of Science, Lanzhou 730000, China;  
2. University of Chinese Academy of Science, Beijing 100049, China)

**Abstract:** It combined daily precipitation data of national Meteorological stations Chinese inland rivers and lakes hydrological stations in Shiyang River Basin in 2010—2012, with DEM data to establish a partial least squares (PLS) regression model between precipitation and geography factors including latitude, longitude, height, slope and aspect which extracted from DEM data. Then, the precipitation distribution characteristic in Shiyang River Basin are presented in annual, seasonal and monthly scale, the results showed that: 1) using the cross-validation to verify the accuracy of the PLS, IDW, Spline, ordinary kriging interpolation methods with the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) value, PLS has the lowest values. And except for the fitting correlation coefficient of precipitation in winter is 0.68 and 0.66 in 2011/2012 year respectively, the other models all above 0.81, with significantly more than 95%. 2) Different time and time scale, the influence of terrain factors for precipitation is different. Basically, altitude and slope have larger influence on precipitation, and the influence is positive. 3) The annual precipitation of Shiyang River Basin from 2010—2012 decreased from the south to the north area, and the maximum occurred in southern Qilian Mountains; The magnitude of seasonal precipitation sequentially for summer > autumn > spring > winter; The precipitation characteristics in 6/7/8 months of different years are also different, in July 2011 and August 2012, precipitation is relatively richer.

**Key words:** Shiyang River Basin; DEM; PLS; Precipitation