

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. 地球物理学报, 2007, 50(6):1703~1712  
Shu S.J., Wang Y., Xiong A.Y. Estimation and analysis for geographic and orographic influences on precipitation distribution in China. *Chinese J. Geophys.* (in Chinese), 2007, 50(6):1703~1712

# 中国区域地理、地形因子对降水分布影响的估算和分析

舒守娟<sup>1</sup>,王元<sup>1\*</sup>,熊安元<sup>2</sup>

1 南京大学大气科学系中尺度灾害性天气教育部重点实验室,南京 210093  
2 国家气象信息中心气象资料室,北京 100081

**摘 要** 不同与以往基于最小二乘的多元线性回归方法,本文首次尝试将新型的二代回归分析方法——偏最小二乘回归分析方法应用到中国区域的降水建模中.利用区域内 394 个气象观测站建站到 2000 年 45 年(及以上)的降水资料,建立了一个简单的年、季降水量和地理、地形因子(包括纬度、经度、地形高程、坡度、坡向和遮蔽度)的关系模型,估算了区域降水量中地理、地形的影响部分,并分析了这种影响的特征.结果表明,用此方法建立的模型能够解释 70% 以上的因变量的变异,相关系数基本都在 0.84 以上,经交叉有效性检验,模型的回归效果较显著.分析表明,在多元线性回归不适用的情况下,本文基于偏最小二乘法的简单模型能够比较准确地定性、定量地再现实际降水分布.

**关键词** 降水空间分布,地形因子,地理因子,偏最小二乘回归

文章编号 0001-5733(2007)06-1703-10

中图分类号 P426

收稿日期 2006-12-11,2007-08-05 收修定稿

## Estimation and analysis for geographic and orographic influences on precipitation distribution in China

SHU Shou-Juan<sup>1</sup>, WANG Yuan<sup>1\*</sup>, XIONG An-Yuan<sup>2</sup>

1 Key Laboratory of Mesoscale Severe Weather of Ministry of Education, Department of Atmospheric Sciences, Nanjing University, Nanjing 210093, China

2 Climatic Data Center, National Meteorological Information Center, China Meteorological Administration, Beijing 100081, China

**Abstract** Different from the framework of multiple linear regression based on least square method, this paper tries to apply a novel second-generation regression method based on partial least-squares to precipitation estimation in China for the first time. The 45-yr precipitation data from 394 meteorological stations in study area are used. Several simple formulae used to estimate the annual mean and seasonal precipitation have been obtained, and the characteristics of the geographic or topographic effects have been presented. The impact factors include longitude, latitude, height, slope, sloping direction and close limit. The results show that the fraction of the variation of response explained by the model is above 70%, and the average correlation coefficients are nearly all above 0.84. The results are satisfied through the test of cross-validation. Through it is not appropriate to set up multiple linear regression models, the estimated precipitation based on partial least-squares regression correctly replicates real spatial distribution of precipitation qualitatively and quantitatively.

**Keywords** Spatial distribution of precipitation, Orographic factor, Geographic factor, Partial least-squares regression

**基金项目** 国家科技基础条件平台建设项目(2005DKA31700)、国家自然科学基金(40575017)、国家重点基础研究发展规划项目(973:2004CB418301)和江苏省自然科学基金项目(BK2005081)资助.

**作者简介** 舒守娟,女,1980年生,现为南京大学大气科学系讲师,主要从事中尺度天气动力学方面的研究.

\*通讯作者 王元,男,南京大学教授、博士生导师. E-mail: yuansm@nju.edu.cn

## 1 引言

我国位于世界最大的大陆——东亚大陆的东部,面向世界最大的大洋——太平洋,自青藏高原向东到太平洋,西高东低、三级阶梯逐级递减是我国地形的主要特征(见图 1a).独特的地理位置和地势形成了我国复杂而多变的降水气候.近年来随着地球科学及交叉学科的发展,对降水量空间分布作分析和估算具有越来越重要的意义和价值.

传统上,估算气象要素的空间分布及变化有两种基本方法<sup>[1,2]</sup>:用内插法来插值现有的观测数据形成等值线的空间分布,但当观测资料稀疏、地形比较复杂时,该方法的准确性很难保证,且结果随插值方法的不同而不同<sup>[3-5]</sup>,其根本原因在于内插法不能充分考虑气象要素空间分布与诸多地理环境要素间的复杂函数关系;用回归分析法来建立气象要素与空间变量如地理位置、地形等的关系方程,这种方法在定量反映真实地形上各要素的空间分布时具有优势.

降水的空间分布取决于如地理、地形因素,以及诸如气团远离水汽源的距离<sup>[6]</sup>等气象条件.由于单纯考虑地理地形条件影响下的降水即地形性降水的分布,其影响是相对常定的,因此,不少学者使用基于静态最小二乘法的多元回归分析方法建立降水与地理地形等影响因子间的关系,推算降水的空间变化.如 Ollinger 等<sup>[1]</sup>建立了北美地区降水等气象要素与地理位置、地形高程等的回归模型,Naoum 等<sup>[6]</sup>给出了地形性降水的多元回归关系,Kravchenko 等<sup>[7]</sup>用地理要素统计方法探讨了年平均降水的推算,Sevruk 等<sup>[8]</sup>考虑了坡度、坡向和地形高程对小流域降水分布的影响,Germann 等<sup>[9]</sup>用降水观测值与地形的关系建立了精度较高的统计分布模式,Wolting<sup>[10]</sup>、Marqu ́ez 等<sup>[11]</sup>也都深入探讨过多地形变量估算降水分布的问题.在国内,傅抱璞<sup>[12]</sup>及其他学者<sup>[13-18]</sup>也有过类似的研究.上述研究表明,多元统计回归法在研究范围较小、影响降水分布的因子明确且相互独立时效果较好,但对较大范围的降水空间分布却难以给出准确的估算和清晰的分析结果<sup>[17]</sup>,其主要原因就是基于普通最小二乘法 (Ordinary Least-Squares, 下文简称 OLS) 的多元统计回归方法在自变量之间存在多重相关性(指自变量之间存在线性相关的现象)时,很难客观地在建模过程中选择恰当的自变量<sup>[19,20]</sup>,使得模型不能包含所

有对因变量有重要解释意义的因素,因而降低了建模的精确性和可靠性.

为建立一个精确可靠的统计分布模型,本文设计了基于偏最小二乘法 (Partial Least-Squares, 下文简称 PLS) 的回归分析方法<sup>[19,20]</sup>.被誉为第二代统计回归分析法的 PLS 回归方法自 1966 年由 Wold<sup>[21]</sup>首次提出以来,近十几年来,因其应用上的便捷性(如能有效克服变量多重相关性、样本数目或多或少不受限制等)、出色的预测能力及能实现多种数据分析方法的综合应用(多元线性回归、典型相关分析、主成分分析)在化学、生物等众多领域得到飞速发展和有效应用<sup>[22,23]</sup>,近年来已经开始应用到气象科学领域<sup>[24]</sup>.

## 2 资料来源和处理

研究区域为 98°E ~ 133°E, 17°N ~ 55°N 范围(图 1, 下文中将简称中国区域).利用中国气象局多年逐日降水整编资料,选取区域内共 394 个典型观测站点(图 1a 中白色“+”号所示)作为样本点,按照地理位置分为七大区域即:西北、西南、华东、华北、东北、华中以及华南区(图 1b)后,计算站点记录达 45 年及以上自建站到 2000 年的年、季平均降水量.对年平均降水量,缺测 20d 以上的该年份不予考虑;对季平均降水量,缺测 7d 以上的该年份不予考虑.所有降水量均包括由雨夹雪、米雪、雾、露、冰针等引起的微量降水.

为避免过大的计算量,本文采用全球卫星遥感 (NOAA/Digit/Data Sets of Terrain 1996) 10 水平分辨率的数字化高程地形资料,同文献<sup>[13]</sup>中各个地形因子同样的定义和做法,同时使用李占清等<sup>[25]</sup>坡度、坡向的分析表达式(设山区地形高程为  $h(x, y)$ ):

$$= \arctan \left[ \left( \frac{\partial h}{\partial x} \right)^2 + \left( \frac{\partial h}{\partial y} \right)^2 \right]^{1/2}, \quad (1)$$

$$= - \arctan \left[ \frac{\partial h / \partial x}{\partial h / \partial y} \right], \quad (2)$$

得到各格点的高程、经度、纬度及坡度、坡向和遮蔽度.遮蔽度是计算 100 km 半径的圆内,格点高程大于中心点高程的网格点数所占的比例.

## 3 研究方法和原理

### 3.1 统计分析

研究区域面积广阔,西高东低的地形特点,十分

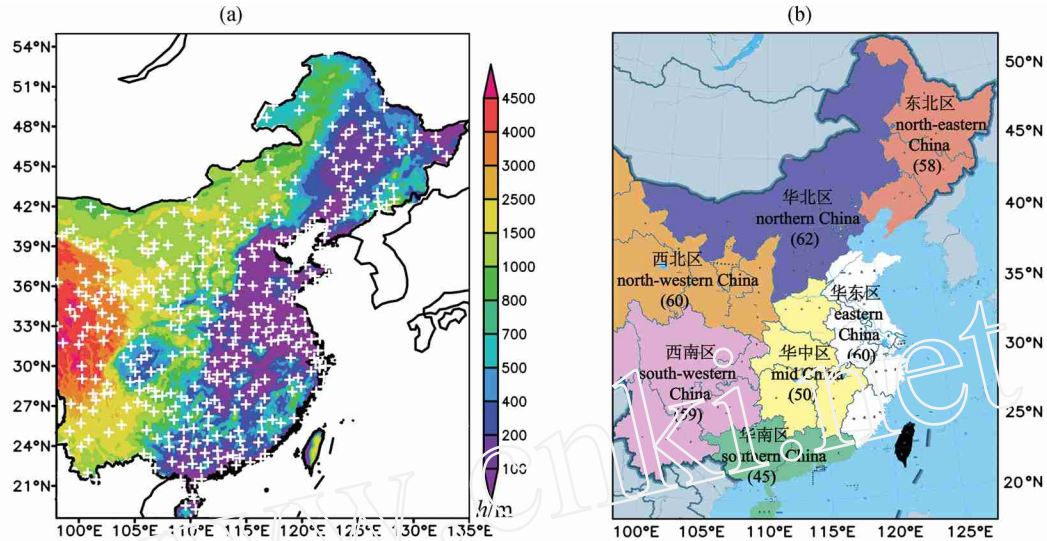


图 1 研究区域概况

(a) 地形高程及站点分布;(b) 政区分布。(a) 中  $h$  表示地形,即图中彩色阴影部分,白色加号表示站点(共 394 个);(b) 中括号内数值表示位于该区域的站点总数。

Fig. 1 The general situation for study area

(a) Distribution of satellite digit elevation and the stations; (b) The districts distribution. In (a),  $h$  is topography and represented by color shading, white plus symbols represent weather stations (the total number is 394); in (b), the values in parentheses represent the number of stations distributed in that region.

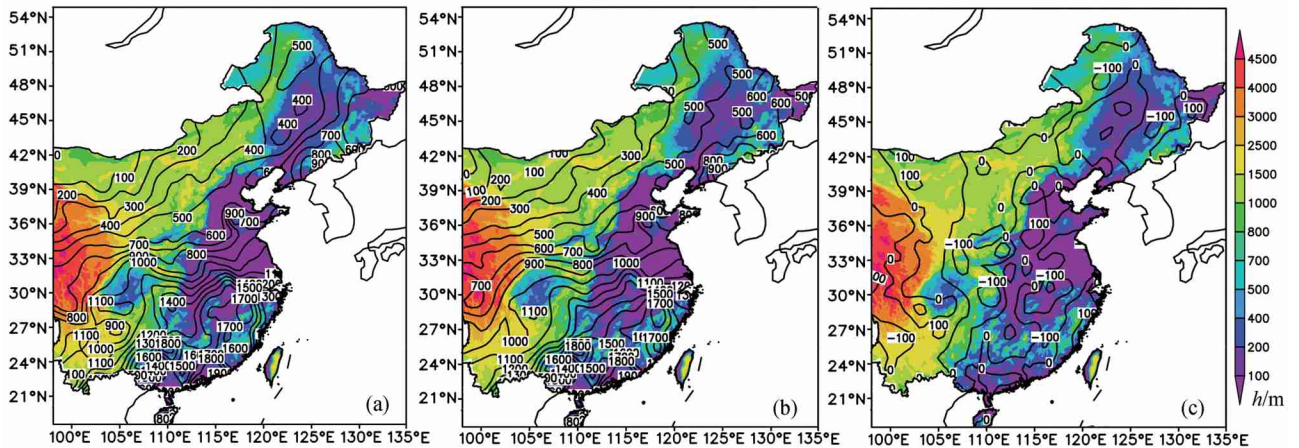


图 3 研究区域实际年平均降水(a),估算年平均降水(b)和两者的误差分布(c)

$h$  表示地形,即图中彩色阴影部分,等值线代表降水或误差,单位为 mm,间隔为 100 mm。

Fig. 3 Distributions for annual mean precipitation observation (a) and its estimation (b),

as well as the estimating error between them (c) in the study area

$h$  is topography and represented by color shading, contours represent precipitation or its error with an interval of 100 mm.

有利于东、南方向的暖湿气流深入内地,气流受到的扰动和阻滞作用使得区域内降水与地形高程、坡度、坡向等形成复杂的相关关系。为了揭示这种关系并估算空间各点上的降水量,本文采用 ELS 统计方法,建立地理、地形因子与降水分布的关系,并认为这是地理、地形影响产生的降水部分<sup>[12]</sup>。在一个较大地区范围内,降水量  $P = P(\phi, \lambda, h, \alpha, \beta, k, e)$ , 其中

$\phi$ , 分别为纬度和经度; $h$  为地形高程; $\alpha$ , 分别为坡度和坡向; $k$  为遮蔽度; $e$  为影响降水的其他因素,并用回归分析中的残差表示,则地理、地形影响下的估算降水量为

$$P = a_0 + a_1 \phi + a_2 \lambda + a_3 h + a_4 \alpha + a_5 \beta + a_6 k + e, \quad (3)$$

式中  $a_0$  为常数项,  $a_1 \sim a_6$  是各项的系数。估算降水

量的水平分辨率也为 10 ,同地形高程的分辨率.

需要指出的是,上述表达式中的 6 个自变量可以分为两类,即地形性因子组分  $h, \dots, k$  和地理因子组分  $\phi, \dots$ ;后者事实上能够反映大的气候和环流背景,这种考虑显然与中国处于东亚季风影响区域相关.

### 3.2 偏最小二乘(PLS)回归方法

#### 3.2.1 多重相关性分析

如引言所述,常用的多元统计回归法在估算大区域降水空间分布时,计算值与实测值存在较大误差<sup>[17]</sup>,究其原因不外乎两个:自变量系统的解释性不够;其次,自变量之间的多重相关性使得某些重要影响因子不被引入.本研究区域经度横跨 35°,西高东低的地势间接造就了地理、地形诸因子间的多重相关性.下文举例说明在研究区域内,用基于 OLS 的多元线性回归建模的不可行性.

表 1 为西南地区 6 个地理、地形因子(作为自变量)和春季降水(因变量)的相关系数(括号内值为西北地区自变量和冬季降水的相关系数),表中各变量

意义同前.可见,自变量之间存在着多重相关性, $r(\phi, h) = -0.73, r(\phi, \dots) = -0.37$ .基于 OLS 建立回归模型后,模型通过 F 检验,但观察回归系数、t 检验值以及回归系数显著为零的概率(SigT)表 2,则显示了多重相关性的典型病兆,因子  $h$  和  $\dots$  均未通过 t 检验,它们显著为零的概率都在 50%左右,表明建模时应剔除,而从相关系数表上看, $h$  与因变量的相关程度却高达 -0.66,对海拔较高的西南地区来说,高程显然是影响降水分布的一个重要因子,如果剔除,其结果很难令人信服.同样,对西北地区冬季降水来说,竟有大半的因子未通过 t 检验(表 2 中括号内数值),显著为零的概率高达 88%,而它与因变量的相关系数却达 0.49,如果剔除则不能准确反映西北地区的降水分布.可见,自变量的多重相关性使得某些对因变量有重要解释意义的因素未能被包含在 OLS 的回归模型中,这势必影响了模型的精确性和稳健性,因此,本文设计了能有效克服变量多重相关性的 PLS 回归分析建模方法.

表 1 西南地区春季降水与(括号内为西北地区冬季降水,表 2 同)的各变量相关系数

Table 1 Correlation coefficients between variables for spring precipitation in south western China and winter precipitation in north western China (in parentheses, the same in Table 2)

$r(\cdot)$	$h$	$\dots$	$k$	$P$		
$\phi$	0.03(-0.21)	0.32(-0.22)	-0.09(-0.04)	-0.01(0.23)	0.21(0.23)	-0.22(-0.77)
		-0.73(-0.72)	-0.37(-0.15)	-0.09(-0.29)	0.05(-0.16)	0.81(0.49)
$h$			0.30(0.02)	0.16(0.09)	0.16(0.05)	-0.66(-0.23)
				0.01(0.29)	0.13(-0.30)	-0.15(-0.08)
					0.25(0.06)	-0.08(-0.19)
$k$						-0.18(-0.13)

表 2 最小二乘法回归建模的系数及其 t 检验

Table 2 Coefficient and its t test based on least squares regression model

变量	回归系数	t 检验值	SigT
$\phi$	-0.204(-0.903)	-2.678(-11.261)	0.01(0.00)
	0.960(0.018)	8.859(0.156)	0.00(0.88)
$h$	0.076(-0.429)	0.671(-3.840)	0.51(0.00)
	0.197(-0.099)	2.751(-1.375)	0.01(0.18)
	0.051(0.086)	0.755(1.222)	0.47(0.23)
$k$	-0.237(0.075)	-3.338(1.076)	0.00(0.29)

#### 3.2.2 PLS 建模

回归建模的目的是通过分析观测数据建立统计模型,辨识数据中的系统因素,并将它与随机误差区别开.为达到  $Y = BX + \dots$  的建模目的( $Y$  和  $X$  分别是因变量和自变量向量, $B$  是系数向量, $\dots$  是随机误差向量),多元线性回归法首先假定  $Y$  估计量  $\hat{Y} =$

$BX$ ,再在  $\hat{Y}$  与  $Y$  的差异即  $e_{LS} = Y - \hat{Y}$  最小的约束下求得参数的最小二乘估计量  $B_{LS}$ ,得到  $\hat{Y} = B_{LS} X + e_{LS}$ ,最后进行 F 检验和 t 检验,完成多元 OLS 回归建模(见流程图 2 左半边).

多元 PLS 回归建模则吸收了主成分分析的精髓,采取先提取成分的方式,如图 2 右半边所示,首先从自变量向量  $X$  中提取成分  $t_1$ ,它既是自变量的线性组合也能最大程度地携带  $X$  中的数据变异,假定估计量  $\hat{Y} = B t_1$ ,再在  $t_1$  对  $Y$  有最大解释能力即两者相关度达最大的约束下实施  $Y$  对  $t_1$  的回归,得到  $\hat{Y} = B_{PLS1} t_1 + e_Y$ ,最后进行交叉有效性检验,如果方程没有达到满意的精度,则利用  $X$  和  $Y$  被  $t_1$  解释后的残余信息  $e_X$  和  $e_Y$  进行第二轮成分提取,直到达到满意精度.最终,若通过交叉有效性检验的成分个数为  $m$ ,则估计量  $\hat{Y} = \sum_{i=1}^m B_{PLSi} t_i$ ,再还原成关于

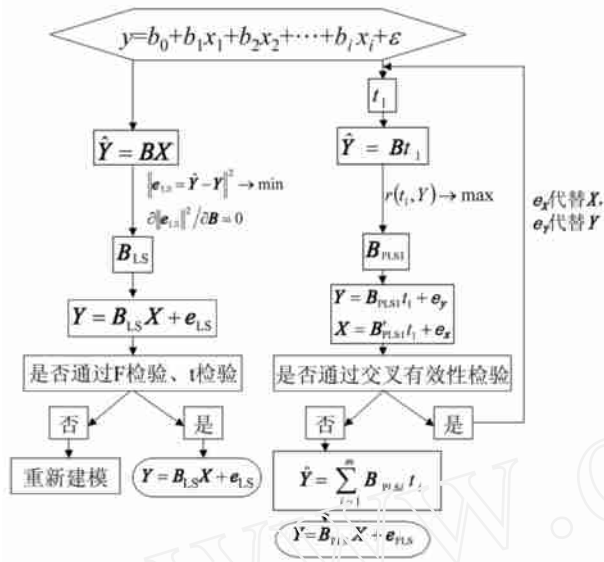


图2 多元 OLS 回归建模(左)和 PLS 回归建模(右)的流程图

Fig.2 Flow chart for multiple OLS regression modeling (left) and PLS regression modeling (right)

原始变量的回归方程  $Y = B_{OLS} X + e_{OLS}$ , 完成 PLS 回归的建模。

PLS 回归分析所采用的交叉有效性判据是这样规定的. 对  $n$  个样本, 记  $y_i$  为因变量原始数据,  $m$  是 PLS 回归过程中提取的成分个数.  $\hat{y}_{mi}$  是使用全部样本点并取  $m$  个成分回归建模后, 第  $i$  个样本点的拟合值.  $\hat{y}_{m(-i)}$  是在建模时删去样本点  $i$ , 取  $m$  个成分回归建模后, 再用此模型计算的  $y_i$  的拟合值. 记  $y_i$  的预测误差平方和为  $^{pre}SS_m = \sum_{i=1}^n (y_i - \hat{y}_{m(-i)})^2$ ,  $y_i$  的误差平方和为  $SS_m = \sum_{i=1}^n (y_i - \hat{y}_{mi})^2$ . 由于  $^{pre}SS_m$  同时含有样本点的扰动误差和拟合方程的预测误差, 因此一般情况下, 总有  $^{pre}SS_m$  大于  $SS_m$  (只含有拟合方程的预测误差),  $SS_m$  小于  $SS_{m-1}$  (用全部样本点拟合的具有  $m-1$  个成分的方程的拟合误差), 即提取成分越多, 模型的解释能力越强. 增加新的成分对模型预测是否有益, 是根据新成分  $t_m$  的交叉有效性  $Q_m^2$  来判断的:

$$Q_m^2 = 1 - \frac{^{pre}SS_m}{SS_{m-1}} \quad (4)$$

如果  $\sqrt{^{pre}SS_m} < 0.95 \sqrt{SS_{m-1}}$ , 则表明增加成分  $t_m$  是有益的, 即当  $Q_m^2 > 0.0975$  时, 引进新的成分会明显改善模型的预测能力, 否则不引进新成分. 可见, 通过交叉有效性检验的各成分的累积交叉有效性实质

性地体现了模型对因变量的总的预测能力.

### 3.3 降水分布模型

利用方程(3), 本文设计了 PLS 回归建模和检验方法, 得到中国区域七个地区春(3~5月)、夏(6~8月)、秋(9~11月)、冬(12~次年2月)四个季节及年平均降水空间的分布估算模型, 如表3(表中各变量意义同前)所示. 表中已略去三个回归效果较差如华中地区夏季、华南地区年及夏季的模型, 这反映出在上述地区特殊季节里, 地理、地形对降水分布的影响并不显著, 估算降水必须重新确定自变量系统(我们将另文讨论), 其他地区都取得了较好的建模效果. 由表中各系数不但可以得到各区域的降水空间分布, 而且其正负值还表征各因子对地形性降水的贡献多少, 以达到实际应用的目的.

## 4 结果检验与分析

### 4.1 结果检验

表4给出了各个区域 PLS 回归建模的精度评价, 其中  $R_Y$  是建模所取自变量成分对因变量总的解释能力, 即模型预测和实际降水量间相关系数的平方;  $Q^2$  则体现由交叉有效性(即方程(4))判断回归模型对降水总的预测能力. 对  $Q^2 < 0.36$  的模型用短实线来代替, 此时较低的  $Q^2$  表示模型的解释和预测能力较差, 也即表3中略去的几个封闭模型. 由表可见, 大多数模型能够解释70%以上因变量的变异, 使得模型的预测和实际降水的相关系数达到0.84(即  $\sqrt{0.7}$ )以上, 最高如西北地区达到了0.94. 各个模型对降水的预测能力都通过了交叉有效性检验, 其中西北地区达到了82%, 预测能力较显著. 相比之下, 西南地区的建模效果较差, 这是因为该区范围最广, 地形最为复杂, 在各种因素的复杂相互作用下, 地理、地形影响下的降水特征很难清晰体现出来. 另外, 利用 PLS 回归模型估算了中国区域年平均降水的空间分布(图3a)(为了完整性, 华南地区年平均降水已用实际值代替), 它与实际降水(图3b)空间分布比较一致. 较大的估算误差(图3c)主要分布在地形复杂、站点分布稀疏, 降水信息量不足的局部地区. 季节性降水的空间分布结果类似(图略).

### 4.2 结果分析

通过观察和分析标准化回归系数直方图, 可以直观地看出每一个自变量在解释因变量集合时的重要性的和它们的相关程度, 以便比较几类地区的降水特征.

表 3 中国区域的降水估算模型

Table 3 The regression models for precipitation estimation in China

		$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$
西北	年	1618.08	- 81.31	15.84	0.01	- 6.69	- 0.25	268.19
	春	419.80	- 18.84	3.15	0.00	1.49	- 0.05	55.17
	夏	712.83	- 35.73	7.17	0.02	- 7.24	- 0.14	122.22
	秋	458.20	- 25.11	5.00	0.00	0.81	- 0.06	107.06
	冬	53.82	- 2.54	0.44	0.00	- 0.50	0.00	6.66
西南	年	687.92	- 34.40	15.84	- 0.07	38.68	- 0.01	- 576.50
	春	- 1655.08	- 3.92	19.64	- 0.02	38.08	0.07	- 164.35
	夏	1187.69	- 16.75	0.23	- 0.03	- 4.44	- 0.07	- 349.52
	秋	162.45	- 5.71	2.89	- 0.02	5.52	- 0.03	- 89.76
	冬	- 152.32	- 2.48	2.88	- 0.01	- 4.16	- 0.01	- 32.94
华东	年	- 845.34	- 70.59	31.64	0.18	305.32	- 0.09	757.65
	春	171.24	- 37.58	10.25	0.02	216.04	- 0.11	159.45
	夏	1185.90	- 6.47	- 5.54	0.09	65.34	- 0.01	308.56
	秋	- 1481.34	- 12.08	16.43	0.03	- 13.39	0.06	215.78
	冬	- 655.75	- 14.46	9.64	0.00	52.61	0.01	133.58
华北	年	21.59	- 18.90	14.79	- 0.16	120.66	- 0.21	- 839.42
	春	82.93	- 3.64	1.67	- 0.02	29.21	- 0.04	- 122.35
	夏	- 213.76	- 9.99	10.86	- 0.11	44.57	- 0.12	- 515.83
	秋	172.49	- 4.75	1.81	- 0.03	37.73	- 0.04	- 197.65
	冬	10.34	- 0.45	0.24	0.00	2.68	0.00	- 16.59
东北	年	1169.84	- 15.00	2.70	0.20	322.45	- 0.08	- 691.12
	春	88.74	- 2.93	1.45	0.06	68.65	- 0.01	- 161.45
	夏	1015.18	- 8.90	- 0.96	0.10	164.70	- 0.06	- 317.29
	秋	69.70	- 2.63	1.75	0.03	68.07	- 0.01	- 166.95
	冬	0.54	- 0.57	0.45	0.02	21.18	0.00	- 48.04
华中	年	- 327.36	- 101.76	37.22	- 0.48	184.04	- 0.17	899.93
	春	- 541.66	- 59.10	22.70	- 0.27	90.45	- 0.10	396.06
	秋	881.34	- 6.39	- 4.77	0.02	27.06	0.03	85.08
	冬	- 273.46	- 22.05	8.97	- 0.11	33.33	0.00	113.76
华南	春	- 2299.44	34.82	17.74	- 0.18	205.37	- 0.17	- 11.63
	秋	1483.39	- 39.39	2.92	- 0.08	4.45	0.05	- 1266.78
	冬	- 514.48	10.66	4.05	- 0.09	69.17	- 0.02	- 116.38

表 4 各区域偏最小二乘回归建模的精度评价

Table 4 The precision assessment for the model based on the partial least-squares regress method

		西北	西南	华东	华北	东北	华中	华南
年	$RY(R^2)$	0.93	0.90	0.96	0.89	0.73	0.91	—
	$Q^2$	0.89	0.84	0.94	0.84	0.68	0.86	—
春季	$RY(R^2)$	0.92	0.73	0.96	0.82	0.75	0.92	0.74
	$Q^2$	0.87	0.65	0.93	0.68	0.70	0.89	0.67
夏季	$RY(R^2)$	0.88	0.58	0.51	0.89	0.68	—	—
	$Q^2$	0.82	0.46	0.42	0.83	0.63	—	—
秋季	$RY(R^2)$	0.92	0.52	0.78	0.76	0.64	0.58	0.85
	$Q^2$	0.87	0.46	0.68	0.60	0.59	0.57	0.80
冬季	$RY(R^2)$	0.75	0.53	0.94	0.50	0.70	0.93	0.62
	$Q^2$	0.66	0.49	0.92	0.37	0.63	0.90	0.49
平均	$RY(R^2)$	0.88	0.65	0.83	0.77	0.70	0.84	0.74
	$Q^2$	0.82	0.58	0.78	0.66	0.65	0.81	0.65

如图 4a 所示,对西北地区的降水来说,纬度变化是其最重要的影响因素,且越往北降水越少。其次,经度影响也较显著,越往东降水越多,这说明我国西北地区(新疆除外)的降水主要受区域性地理因子的影响,区域差异较大,水汽的南北输送最显著,东西输送次之,年平均和季节降水的特征差别不大。但在夏季,高程也是影响降水的一个正相关因子。

西南地区的降水在各个季节呈现出不同的特征(图 4b),体现出季风变化下地理地形因素的复杂影响。春、冬两季,经度和地形高程是影响降水分布的首要因子。此时,西太平洋上的水汽在东北季风的影响下能被输送到西南地区<sup>[13]</sup>,降水分布的东西差异较大,越向西降水越少,且有山脉阻挡的地方降水更少。而在夏季,西南季风将暖湿空气带到该区,纬度和地形高程的负相关作用最显著,降水的南北差异

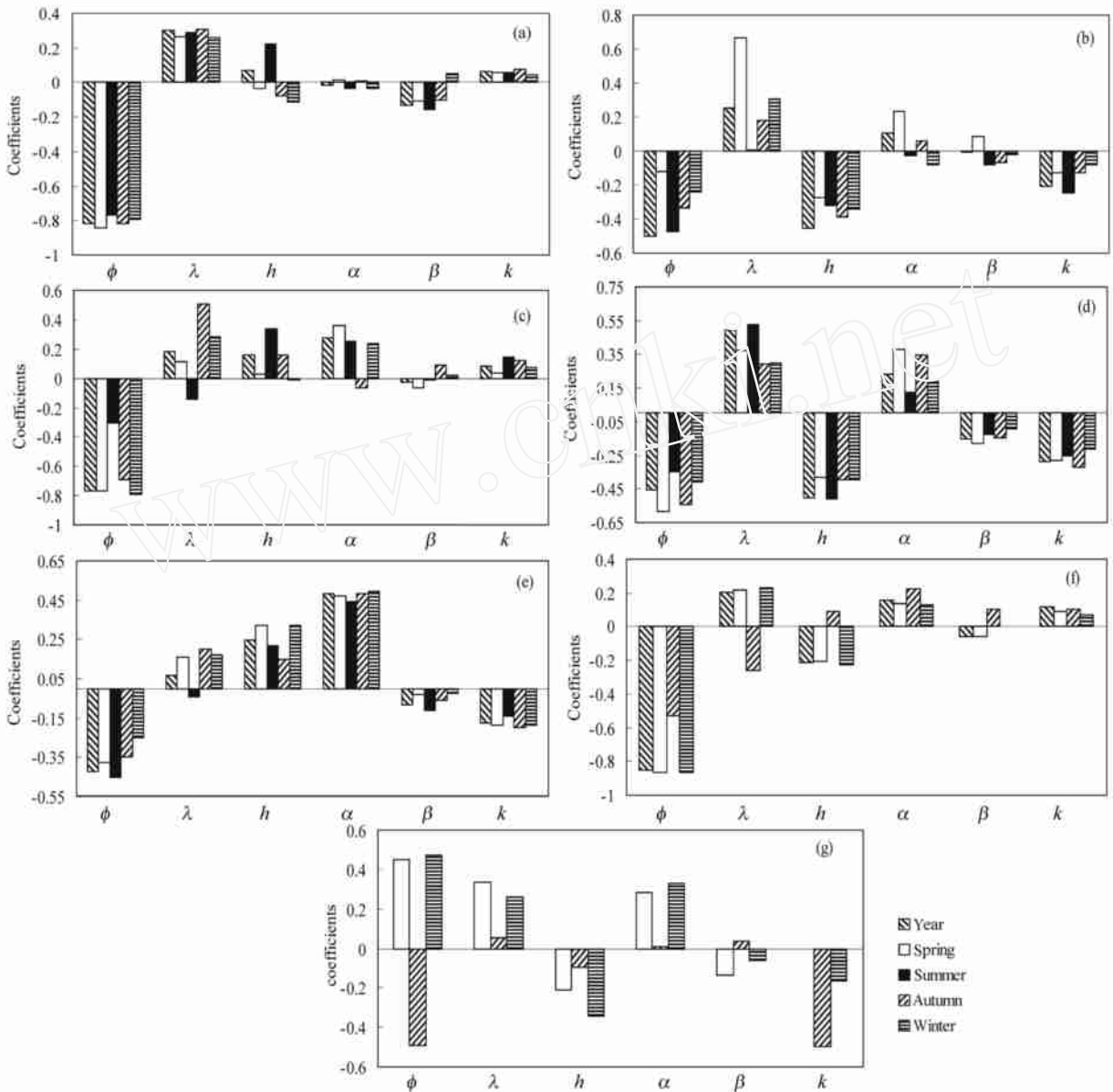


图 4 各个地区降水的标准化回归系数直方图

(a) 西北地区;(b) 西南地区;(c) 华东地区;(d) 华北地区;(e) 东北地区;(f) 华中地区;(g) 华南地区.

Fig. 4 Standard regression coefficients histograms for precipitation in (a) north-western China , (b) south-western China , (c) eastern China , (d) northern China , (e) north-eastern China , (f) mid China and (g) southern China

较大,越往北降水越少,年平均、秋季的降水特征同夏季类似.

华东地区夏季降水分布同时受到区域地理和局地地形因子的影响,各因子的影响错综复杂.而在其他季节,主要是区域性地理因子纬度、经度的变化造成了降水的空间差异,尤其是纬度变化的影响.华东地区位于我国东部太平洋沿岸,地势较平坦,降水主要是在大型环流背景之下形成,区域性差异较大,如越往北或西降水越少.但在夏季,受多尺度系统如西太平洋台风等的影响,降水分布还会受到区域因子

以外的个别局地地形因子(高程、坡度等)的影响(图 4c),地形高程和坡度的增大都有利于降水的增加.

从图 4d 可以看出,华北地区回归模型的系数结构非常一致,说明该区各个季节降水的影响因子类似,各季降水特征类似.该区环流形势较为简单,在冬、夏季风环流的影响下,降水总是受到区域性地理因子经度、纬度和局地地形高程的影响,越往东或南降水越多,并且和高程呈反相关关系.

由图 4e 可见,东北地区有非常明显的局地地形降水,坡度和高程是影响降水分布的主要正相关因

子,降水同时也受到区域因子纬度的负相关影响,各个季节的降水特征比较类似.东北地区的地形特征是东、北、西三面为山环绕,中间为广阔的大平原,区域地形的海拔虽然不高,但坡度变化比较明显.受环流和地形的综合影响,东北地区全年都有较大范围东南、西南或南风的盛行分布<sup>[26]</sup>.在坡度和高程的正相关作用下,长白山、大、小兴安岭对这些盛行风向的屏障作用形成山区丰富的降水,而平原地区的降水很少,且降水都是往北、往西而减少,因此长白山地的东南侧(也即迎风坡)是本区降水最多的地方,这些结论与文献<sup>[26,27]</sup>的总结相当一致.

由图 4f 可见,华中地区降水的显著特征是降水与区域地理因子——纬度的负相关性非常大,其他因子的影响都很微弱,降水的季节差异不大,但却有着较大的南北区域性差异,越往北降水越少,这可能暗示着影响该区的降水系统主要是南北向移动系统.

华南地区的降水特征复杂多变,且季节性差异较大(图 4g),这与当地所受天气系统复杂多变是分不开的.在春季,孟加拉湾低槽前的西南气流控制中南半岛,副热带高压控制我国南海地区<sup>[28]</sup>.西南气流与副高西侧偏南气流汇合形成一股强劲的暖湿风带,当与移到华南地区的北方变性冷气团相遇时,常造成该区春季持续性降水.从图 4g 可见,在这种大型环流形势影响下,华南地区春季降水明显受到区域性地理因子纬度、经度的影响,较多的降水形成在华南地区靠北、东的地方.秋季降水越往北越少,受到的遮蔽程度越大降水越少,可见水汽应主要来源于南部洋面上.冬季降水主要在寒潮南下过程中形

成,越靠北降水越多,遇到有利地形时,坡度的增加会增加降水.华南地区是惟一纬度因子起正相关作用(春、冬两季)的区域,此时降水都与北方南下冷空气运动过程有关.夏季,环境流场时有突变<sup>[28]</sup>,副高东西摆动和北跳,使得华南地区交替出现西南、东南气流;另外南海、西太平洋上的低压扰动系统时有发生,使夏季降水预报变得极其复杂,此时的回归模型没有通过交叉有效性检验,我们将另文讨论.

此外,PLS 回归分析法的另一个优势是它能提供低维空间更直观的有效产品.根据交叉有效性,当模型取一个以上成分时,在第一成分  $t_1$ 、第二成分  $t_2$  所构造的二维空间上,以  $(r(x_j, t_1), r(x_j, t_2))$  绘制自变量  $x_j$ (图 5 中黑点所示),以  $(r(y, t_1), r(y, t_2))$  绘制因变量  $y$ (图 5 中星号所示),可以直观地观察各变量间的相关关系结构,得到与上文相似的结论.以西南地区为例(图 5a),在第一成分方向上自变量  $\phi$  和  $h$  相距最远,揭示了两者的极大负相关关系;其次距离降水  $P$  最远、与其处不同象限的  $\phi$ 、 $h$  即为与  $P$  负相关性最大的两个因子.对东北地区而言(图 5b),降水  $P$  与因子  $\lambda$ 、 $h$  靠的最近并处同一象限、与不同象限的  $\phi$ 、 $k$  离的最远,分别揭示了坡度和高程对降水的正相关,以及纬度和遮蔽度对降水的负相关影响;因子  $\alpha$  和  $\beta$  靠近中心点  $(0, 0)$  则表明它们对降水影响的可忽略性.

### 5 结 论

我国地域辽阔,从西部的高原往东,跨越高山、盆地、平原以及低山和丘陵,地势西高东低,降水特

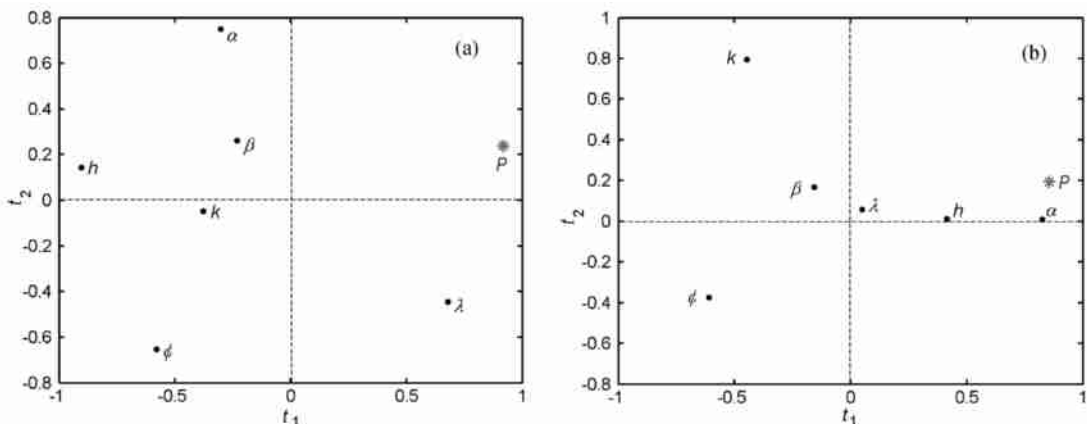


图 5 各变量与  $t_1$ 、 $t_2$  的相关系数图  
(a) 西南地区;(b) 东北地区.

Fig. 5 Correlation coefficient between variable and components of  $t_1$  and  $t_2$  for (a) southwestern and (b) northeastern China



征复杂多样. 经考察, 由于自变量系统的多重相关性, 在研究区域不适合用基于 OLS 的多元线性回归建模的情况下, 本文设计了基于 PLS 的回归分析方法, 使得所有对因变量有重要解释意义的因子都能包含在模型中, 保证了模型的精确性和稳健性. 分析表明, 估算模型能够比较准确地估算我国降水的空间分布, 得出以下主要结论:

在东亚季风环流的影响下, 我国年、季降水量在空间分布上基本呈往北、往西而减少的基本态, 这由文中区域性地理因子纬度、经度对降水之负、正相关影响得到明显体现, 其他学者也有类似的研究结果<sup>[29]</sup>, 但在某些区域的特殊季节, 如在华南地区的春、冬季, 降水随着纬度的增加而增加; 在华中的秋季, 降水越往东越少.

除西北地区以外, 局地地形因子地形高程、坡度及遮蔽度对降水空间分布的影响都起着举足轻重的作用. 东北地区的地形性降水最为显著, 降水随着地形高程和坡度的增加而增加, 在华东地区, 降水也有类似特点, 但在其他区域, 降水随着地形高程的增加而减小. 除华东及华中地区以外, 遮蔽度对降水基本呈负相关的影响.

基于 PLS 回归分析方法, 本文建立的模型能够准确估算我国降水的空间分布, 反映我国降水分布的基本天气、气候学特征, 具有一定的实用价值.

审图号 GS(2007) 1393.

致谢 国家气象信息中心气象资料室为研究工作提供了降水整编资料, 特此致谢.

## 参考文献 (References)

- [ 1 ] Ollinger S V, Aber J D, Federer C A, et al. Modeling physical and chemical climate of the northeastern United States for a geographic information system. Gen. Tech. Rep., 1995. 191
- [ 2 ] Palomino I, Martin F. A simple method for spatial interpolation of the wind in complex terrain. *J. Appl. Meteor.*, 1995, **34**: 1678 ~ 1693
- [ 3 ] Godin W R, Mcrae G J, Seinfeld J H. A comparison of interpolation methods for sparse data: application to wind and concentration field. *J. Appl. Meteor.*, 1979, **18**: 761 ~ 771
- [ 4 ] Bolstad P V, Swift L, Collins F, et al. Measured and predicted air temperatures at basin to regional scales in the southern Appalachian Mountains. *Agric. For. Meteor.*, 1998, **91**: 161 ~ 176
- [ 5 ] Jarvis C H, Stuart N. A comparison among strategies for interpolating maximum and minimum daily air temperatures Part II The interaction between number of guiding variables and the type of interpolation method. *J. Appl. Meteor.*, 2001, **40**: 1075 ~ 1084
- [ 6 ] Naoum S, Tsanis I K. Orographic precipitation modeling with multiple linear regression. *Journal of Hydrologic Engineering*, 2004, **9**: 79 ~ 102
- [ 7 ] Kravchenko A, ZHANG Ren-Duo, Tung Y K. Estimation of Mean Annual Precipitation in Wyoming Using Geostatistical Analysis. Proceedings of American Geophysical Union 16th Annual Hydrology Days, 1996. 271 ~ 282
- [ 8 ] Sevruc B, Nevenic M. The geography and topography effects on the area pattern of precipitation in a small Prealpine Basin. *Wat. Sci. Tech.*, 1998, **37**: 163 ~ 170
- [ 9 ] Germann U, Joss J. Spatial continuity of Alpine precipitation. *Phys. Chem. Earth(B)*, 2000, **25**: 903 ~ 908
- [ 10 ] Wörling G, Bouvier C, Danloux J, et al. Regionalization of extreme precipitation distribution using the principal components of the topographical environment. *Journal of Hydrology*, 2000, **233**: 86 ~ 101
- [ 11 ] Marquinez J, Lastra J, Garcia P. Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis. *Journal of Hydrology*, 2003, **270**: 1 ~ 11
- [ 12 ] 傅抱璞. 地形和海拔高度对降水的影响. *地理学报*, 1992, **47** (4): 303 ~ 314  
Fu B P. The effects of topography and elevation on precipitation. *Acta Geographica Sinica* (in Chinese), 1992, **47**(4): 303 ~ 314
- [ 13 ] 舒守娟, 喻自凤, 王元等. 西藏地区复杂地形下的降水空间分布估算模型. *地球物理学报*, 2005, **48**(3): 535 ~ 542  
Shu S J, Yu Z F, Wang Y, et al. A statistic model for the spatial distribution of precipitation estimation over Tibet complex terrain. *Chinese J. Geophys.* (in Chinese), 2005, **48**(3): 535 ~ 542
- [ 14 ] 张杰, 韩永翔, 万信等. 祁连山区降水资源网格场的模拟与分析. *干旱地区农业研究*, 2002, **20**(2): 108 ~ 119  
Zhang J, Han Y X, Wan X, et al. Simulation and analysis on precipitation resources of gridding field in the Qilian Mountains. *Agricultural Research in the Arid Area* (in Chinese), 2002, **20**(2): 108 ~ 119
- [ 15 ] 张连强, 赵有中, 欧阳宗继等. 运用地理因子推算山区局地降水量的研究. *中国农业气象*, 1996, **17**(2): 6 ~ 10  
Zhang L Q, Zhao Y Z, Ouyang Z J, et al. Studies on local distribution of precipitation in mountains areas calculated with geographic factors. *Chinese Agricultural Meteorology* (in Chinese), 1996, **17**(2): 6 ~ 10
- [ 16 ] 穆兴民, 陈国良. 黄土高原降水与地理因素的空间结构趋势面分析. *干旱区地理*, 1993, **16**(2): 71 ~ 76  
Mu X M, Chen G L. Analysis on the space structure trend area of precipitation and geographical factors in the Loess Plateau. *Arid Land Geography* (in Chinese), 1993, **16**(2): 71 ~ 76
- [ 17 ] 周锁铨, 薛根元, 周丽峰等. 基于 GIS 降水空间分析的逐步插值方法. *气象学报*, 2006, **64**(1): 100 ~ 111  
Zhou S Q, Xue G Y, Zhou L F, et al. The stepwise interpolation approach of precipitation for spatial analysis based on GIS. *Acta Meteorologica Sinica* (in Chinese), 2006, **64**(1): 100 ~ 111
- [ 18 ] 陆忠艳, 马力, 缪启龙等. 起伏地形下重庆降水精细的空间分布. *南京气象学院学报*, 2006, **29**(3): 408 ~ 412  
Lu Z Y, Ma L, Miao Q L, et al. Fine spatial distribution of precipitation on Chongqing rugged terrain. *Journal of Nanjing*

- Institute of Meteorology* (in Chinese), 2006, **29**(3): 408 ~ 412
- [19] 王惠文. 偏最小二乘回归方法及其应用. 北京: 国防工业出版社, 1999. 150 ~ 271  
Wang H W. Partial Least-Squares Regression Method and Applications (in Chinese). Beijing: National Defence Industry Press, 1999. 150 ~ 271
- [20] 任若恩, 王惠文. 多元统计数据分析——理论、方法、实例. 北京: 国防工业出版社, 1997. 149 ~ 163  
Ren R E, Wang H W. Multivariate Statistics and Data Analysis—Theory, Method and Example (in Chinese). Beijing: National Defence Industry Press, 1997. 149 ~ 163
- [21] Wold H. Estimation of principal components and related models by iterative least squares. In: Krishnaiah P R ed. Multivariate Analysis. Academic Press, 1966. 391 ~ 420
- [22] Martens H, Naes T. Multivariate Calibration. New York: Wiley, 1989
- [23] Frank I E, Friedman J H. A statistical view of some chemometrics regression tools (with discussion). *Technometrics*, 1993, **35**: 109 ~ 148
- [24] Wang Yuan. The upper airflow prediction by PLS (Partial Least Square) model. The Scientific Conference on the South China Sea Monsoon Experiment (SCSMEX), Shanghai, China. 2001
- [25] 李占清, 翁笃鸣. 一个计算山地地形参数的计算机模式. 地理学报, 1987, **42**(3): 269 ~ 278  
Li Z Q, Weng D M. A computer model to determine topographic parameters. *Acta Geographica Sinica* (in Chinese), 1987, **42**(3): 269 ~ 278
- [26] 周琳主编. 东北气候. 北京: 气象出版社, 1991. 43 ~ 62  
Zhou L. Climate in Northeast China (in Chinese). Beijing: Meteorological Press, 1991. 43 ~ 62
- [27] 韩艳凤, 江志红, 王冀等. 东北地区夏季降水时空变化特征. 气象科技, 2005, **33**(2): 136 ~ 141  
Han Y F, Jiang Z H, Wang Y, et al. Temporal and spatial characteristics of summer rainfall in Northeast China. *Meteorological Science and Technology* (in Chinese), 2005, **33**(2): 136 ~ 141
- [28] 陈隆勋, 朱乾根, 罗会邦等著. 东亚季风. 北京: 气象出版社, 1991. 68 ~ 72  
Chen L X, Zhu Q G, Luo H B, et al. Monsoon in East Asia (in Chinese). Beijing: Meteorological Press, 1991. 68 ~ 72
- [29] 严华生, 严小冬. 中国降水场的时空分布变化. 云南大学学报(自然科学版), 2004, **26**(1): 38 ~ 43  
Yan H S, Yan X D. The spatial and temporal distribution variation of precipitation field over China. *Journal of Yunnan University (Natural Sciences)* (in Chinese), 2004, **26**(1): 38 ~ 43

(本文编辑 何燕)