

赵中军,刘善亮,游大鸣,等. 偏最小二乘回归模型在辽宁汛期降水预测中的应用[J]. 干旱气象,2015,33(6):1038-1044, [ZHAO Zhongjun, LIU Shanliang, YOU Daming, et al. Application of Partial Least Square Regression on Precipitation Prediction During the Flood Season in Liaoning Province[J]. Journal of Arid Meteorology, 2015, 33(6):1038-1044], doi:10.11755/j.issn.1006-7639(2015)-06-1038

## 偏最小二乘回归模型在辽宁汛期 降水预测中的应用

赵中军<sup>1</sup>,刘善亮<sup>2</sup>,游大鸣<sup>3</sup>,王学忠<sup>3</sup>,胡邦辉<sup>3</sup>,程一帆<sup>4</sup>

(1. 中国人民解放军92493部队,辽宁 葫芦岛 125000;2. 中国人民解放军61741部队,北京 100081;  
3. 解放军理工大学气象海洋学院,江苏 南京 211101;4. 兰州大学大气科学学院,甘肃 兰州 730000)

**摘要:**利用1951~2011年中国160个气象站逐月降水、温度、74项环流指数和NCEP再分析海表温度资料,采用偏最小二乘回归(PLSR)方法,结合均生函数构造预报量周期性因子,建立辽宁省汛期平均降水量及其5站(沈阳、朝阳、营口、丹东和大连)汛期降水量预测模型,并进行预测效果检验分析。结果表明:采用均生函数构造预报量周期性因子,在一定程度上弥补了气候预测统计模型高相关性因子的不足,从而使辽宁汛期平均降水量PLSR模型的试报均方根误差降低约10 mm。PLSR模型由于较好地解决了预报因子之间的多重相关性问题,其预测效果较逐步回归模型有明显提高,对2002~2011年辽宁5站汛期降水量试报的Ps评分平均值为72.6%,比逐步回归模型提高了10.3%。

**关键词:**偏最小二乘回归;均生函数;汛期降水预测

**文章编号:**1006-7639(2015)-06-1038-07 doi:10.11755/j.issn.1006-7639(2015)-06-1038

**中图分类号:**P426.6

**文献标识码:**A

### 引言

辽宁省位于我国东北地区南部、欧亚大陆东岸,属于温带大陆性季风气候<sup>[1]</sup>。辽宁东部和西部为山地丘陵,中部为辽河平原,地形由北向南、自东西两侧向中部倾斜<sup>[2]</sup>。辽宁汛期出现在夏季(6~8月),降水量占全年总降水量的64.8%<sup>[3]</sup>。汛期降水主要受中高纬环流、西太平洋副热带高压、热带环流系统、准2 a振荡(Quasi-Biennial Oscillation, QBO)、太阳活动、ENSO(El Nino La Nina-Southern Oscillation)事件、欧亚遥相关型(Eurasian, EU)等系统影响<sup>[4-6]</sup>。杨文艳<sup>[7]</sup>利用1957~2000年辽宁24个站逐月降水资料,对辽宁汛期降水进行模糊聚类分析,将汛期降水异常分成若干相关性较好的区域,并做了信度检验;李广霞等<sup>[8]</sup>对辽宁夏季降水的基本特征分析发现,辽宁夏季降水存在2~3 a、3~5 a、10~12 a等周期振荡,且暴雨、大暴雨日数空间分布极不均匀。目前,辽宁汛期降水的大多数研究集

中在气候趋势、旱涝影响因子等方面,而对汛期降水预测模型研究相对较少。

多元回归分析是短期气候预测的常用统计学方法,采用最小二乘法估计参数。预报因子之间的较强相关性会严重影响参数的估计,扩大模型误差,并破坏模型的稳定性<sup>[9]</sup>。针对这一问题,1983年Wold等提出了一种新型的多元统计方法——偏最小二乘回归方法(Partial Least Square Regression, PLSR)<sup>[10]</sup>,该方法能够较好地解决预报因子之间的多重相关性问题<sup>[11-13]</sup>。黄颖等<sup>[14]</sup>利用1952~2002年月平均高度、温度和风场资料,采用PLSR方法建立了华南前汛期平均降水量预测模型,预测效果明显好于逐步回归。由于汛期降水不仅受前期大气环流、物理量因子影响,还与预报量自身的周期变化有关<sup>[15]</sup>。因此,本文试图建立辽宁汛期平均降水量的PLSR预测模型,研究预报量周期性因子的作用,在此基础上结合周期因子和常规物理量因子,对其汛期单站降水进行预测,这将对汛期防洪减灾是一种

收稿日期:2015-08-19;改回日期:2015-10-21

基金项目:公益性行业(气象)专项(GYHY201306047, GYHY201206004)资助

作者简介:赵中军(1965-),男,黑龙江五常人,高级工程师,硕士,主要从事水文气象预报技术和保障相关工作。E-mail:zzjHLD@126.com

通讯作者:程一帆(1983-),男,甘肃榆中人,工程师,博士,主要从事现代天气预报技术研究工作。E-mail:chengyf@lzu.edu.cn

很有意义的尝试。

## 1 资料与方法

### 1.1 资料

使用1951年1月至2011年12月的各类资料:(1)国家气候中心提供的160站逐月降水、温度资料及74项环流指数资料;(2)美国国家环境预测中心(NCEP)全球海表温度(SST)再分析资料,水平分辨率 $2^{\circ} \times 2^{\circ}$ 。将资料依照用途分为2部分:1951~2001年的资料用于建模,2002~2011年的资料作为独立样本用于模型检验。

### 1.2 方法

#### 1.2.1 PLSR方法

PLSR方法分多预报量和单预报量2种类型:前一类适合预报因子有限,需要解释多种预报量情形;后一类的预报因子丰富,可专门解释某种预报量。显然后者更适用于汛期降水预测问题。首先,该方法在变量系统中提取若干对系统具有最佳解释能力的新综合变量,称为PLSR成分,然后利用这些成分进行回归建模。考虑单一预报量 $y$ 与 $m$ 个预报因子 $x_1, x_2, \dots, x_m$ 的建模问题。PLSR建模算法如下:

##### (1)数据标准化

预报因子组和预报量的 $n$ 次观测数据的标准化矩阵分别记为 $E_0$ 和 $F_0$ :

$$E_0 = \begin{bmatrix} x_{11}^* & \cdots & x_{1m}^* \\ \vdots & \ddots & \vdots \\ x_{n1}^* & \cdots & x_{nm}^* \end{bmatrix}, \quad F_0 = \begin{bmatrix} y_1^* \\ \vdots \\ y_n^* \end{bmatrix} \quad (1)$$

式(1)中,  $x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j}$ ,  $i = 1, 2, \dots, n; j = 1, 2, \dots, m$ , 表示第 $j$ 个预报因子第 $i$ 次观测的标准化值;  $y_i^* = \frac{y_i - \bar{y}}{s_y}$ ,  $i = 1, 2, \dots, n$ , 表示第 $i$ 次观测对应的预报量标准化值;  $\bar{x}_j, \bar{y}$  分别表示第 $j$ 个预报因子和预报量的平均值,  $s_j, s_y$  分别表示第 $j$ 个预报因子和预报量的标准差。

##### (2)主成分引入

分别提取2变量组的第一对成分,并使之相关性最大。从 $E_0$ 中提取第一成分 $t_1$ ,  $t_1 = E_0 w_1$ ,  $w_1 = \frac{E_0^T F_0}{\|E_0^T F_0\|}$ , 且 $\|w_1\| = 1$ , 实施 $E_0$ 和 $F_0$ 在 $t_1$ 上的回归:

$$\begin{cases} E_0 = t_1 p_1^T + E_1 \\ F_0 = t_1 r_1^T + F_1 \end{cases} \quad (2)$$

式(2)中,  $E_1$ 和 $F_1$ 分别是 $E_0$ 和 $F_0$ 的残差矩阵,  $p_1^T$ 和 $r_1^T$ 是回归系数, 即  $p_1 = \frac{E_0^T t_1}{\|t_1\|^2}$ ,  $r_1 = \frac{F_0^T t_1}{\|t_1\|^2}$ 。用

交叉有效性检验检查 $y$ 对 $t_1$ 回归方程的收敛性,若精度达到要求,则进行下一步;否则,以残差矩阵 $E_1$ 和 $F_1$ 分别代替 $E_0$ 和 $F_0$ ,提取成分 $t_2$ ,循环往复,直到提取了 $h$ 个成分 $t_h$ ,精度达到要求为止。

##### (3)主成分回归

在方程满足精度要求后,得到 $h$ 个成分,实施 $F_0$ 在 $t_1, t_2, \dots, t_h$ 上的回归,得到:

$$F_0 = r_1 t_1 + r_2 t_2 + \cdots + r_h t_h \quad (3)$$

由于 $t_1, t_2, \dots, t_h$ 是 $E_0$ 的线性组合,可得 $F_0 = r_1 E_0 w_1^* + r_2 E_0 w_2^* + \cdots + r_h E_0 w_h^*$ , 其中,  $w_h^* = \prod_{j=1}^{h-1} (I - w_j p_j^T) w_h$ ,  $I$ 为单位矩阵。

##### (4)逆标准化

按照标准化的逆过程,将 $F_0$ 的回归方程还原为 $y$ 对 $x_1, x_2, \dots, x_m$ 的回归方程,即:

$$y = a_0 + a_1 x_1 + \cdots + a_m x_m \quad (4)$$

式(4)中,  $a_0, a_1, \dots, a_m$ 为方程系数。

#### 1.2.2 交叉有效性检验

在PLSR建模中,成分的选取应该通过考察增加一个新成分后,能否对模型的预测功能有明显改进来确定。采用类似于抽样测试法的工作方式,把所有 $n$ 个样本分成2部分,第一部分是除去某个样本点 $i$ 的所有样本点集合(共含 $n-1$ 个样本点),用该集合样本使用 $h$ 个成分拟合一个回归方程;第二部分是将被排除的样本 $i$ 代入前面拟合的回归方程,得到 $y^*$ 在第 $i$ 个样本上的预测值 $\hat{y}_{h(-i)}$ 。对于每一个 $i = 1, 2, \dots, n$ 重复上述测试,则可定义 $y^*$ 的预测误差平方和 $PRESS_h = \sum_{i=1}^n (y_i^* - \hat{y}_{h(-i)})^2$ 。

另外,采用所有样本拟合含 $h$ 个成分的回归方程。记第 $i$ 个样本点的预测值为 $\hat{y}_{hi}^*$ ,则可定义 $y^*$ 的误差平方和 $SS_h = \sum_{i=1}^n (y_i^* - \hat{y}_{hi}^*)^2$ 。

如果 $PRESS_h$ 值小于 $SS_{h-1}$ 时, $h$ 个成分回归方程的扰动误差小于 $(h-1)$ 个成分回归方程的拟合误差,则说明增加一个成分 $t_h$ 会使预报效果提高。为此,定义交叉有效性 $Q_h^2$ ,并取其 $\geq 0.0975$ 作为引入第 $h$ 个成分的条件,即:

$$Q_h^2 = 1 - \frac{PRESS_h}{SS_{h-1}} \geq 0.0975 \quad (5)$$

当满足(5)式时,引入第 $h$ 个成分,并取新 $h$ 的

值为  $h+1$ , 对新  $h$  个成分进行考查; 如此滚动进行, 直至(5)条件得不到满足, 此时  $h-1$  即为引入的主成分个数。

采用均方根误差来检验预报模型效果:  $RMSE =$

$$\sqrt{\sum_{i=1}^n (F_i - O_i)^2 / n}, \text{ 其中 } F_i \text{ 为预测值, } O_i \text{ 为观测值, } n \text{ 为检验样本数。}$$

### 1.2.3 $P_s$ 评分检验

为了检验 PLSR 和逐步回归 (Step Wise Regression, SWR) 2 种方法对辽宁 5 站汛期降雨量的预测效果, 按照中国气象局下发的《短期气候预测质量分级检验方法》进行评分。对各站汛期降雨量预测值按照降水量距平百分率实行 6 级评分制: 特少(1 级)、偏少(2 级)、略少(3 级)、略多(4 级)、偏多(5 级)、特多(6 级)。 $P_s$  评分的计算公式为:

$$P_s = \frac{N_0 + P_1 \times N_1 + P_2 \times N_2}{N + P_1 \times N_1 + P_2 \times N_2} \times 100\% \quad (6)$$

式中,  $P_s$  是汛期降水预测得分;  $N$  是参加评分的总站数;  $N_0$  是预测与实况距平符号相同的站数与符号不

同但误差只有一级的站数之和;  $N_1$  是预测与实况同为偏多、偏少的站数;  $N_2$  是预测与实况同为特多、特少的站数;  $P_1 = 2$  为偏多、偏少级预测成功加分权重系数;  $P_2 = 5$  为特多、特少级预测成功加分权重系数。

## 2 辽宁汛期降水量的变化特征

从辽宁 5 站汛期降水的气候变化特征中看出(图 1), 辽宁汛期 5 站平均降水量具有明显的年代际变化特征, 1958 ~ 1964 年为多雨期, 1967 ~ 1983 年为明显少雨期, 1984 ~ 2011 年为震荡期, 近 61 a 汛期降水呈逐年下降趋势, 但这种下降在 5 站中并不均匀, 沈阳、丹东下降明显, 而朝阳、营口、大连相对平缓。为检验汛期降水变化趋势的显著性, 计算了线性拟合序列与原降水序列的相关系数, 参考自由度为 60 在 0.05 信度水平的临界相关系数为 0.250, 得到沈阳、朝阳、营口、丹东、大连及 5 站平均的相关系数分别为 0.219、0.130、0.066、0.292、0.158、0.252, 可见丹东和 5 站平均的汛期降水在 0.05 信度水平上呈显著减少趋势。

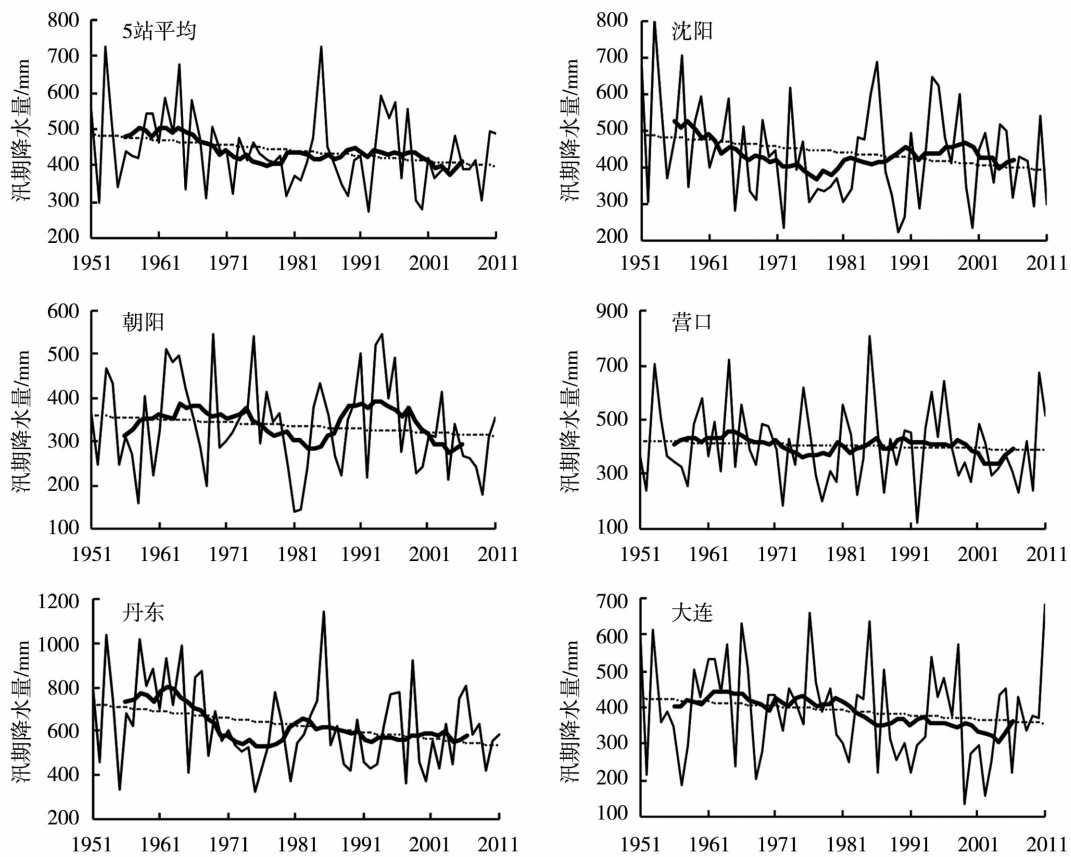


图 1 1951 ~ 2011 年辽宁 5 站汛期降水(细实线)及其 11 a 滑动平均(粗实线)和线性趋势(间断线)

Fig. 1 The changes of precipitation (thin solid line) during the flood season and its 11-year moving average (thick solid line), linear trend (dotted line) from 1951 to 2011 in 5 stations of Liaoning

### 3 辽宁汛期平均降水量 PLSR 预测模型

#### 3.1 预报因子的选取及 PLSR 预测模型

利用 1951 ~ 2001 年数据样本集,当年汛期的预测因子从当年 1 月和 2 月及其前一年 12 个月中的数据中选取,预测对象为当年汛期平均降水量,包含 74 项环流指数(主要代表大气环流特征)、辽宁 5 站平均温度(反映东北低温影响)、SST(代表 ENSO 及其它遥相关的影响),预测对象为当年汛期降水量,求算各因子与辽宁汛期平均降水量(沈阳、朝阳、营口、丹东和大连 5 站汛期降水量的平均值)的相关系数。从相关系数  $> 0.37$  (通过  $\alpha = 0.01$  水平的显著性检验)的因子中,选取 6 个物理意义不同的因子(表 1)。可以看出,6 个因子与汛期平均降水预报量的相关系数均  $> 0.4$ ,其中,大西洋欧洲环流型、亚洲区极涡面积指数、北半球极涡均反映了前期中高纬地区冷空气活动强度;SST 所处位置 ( $54^{\circ}\text{W}$ 、 $66^{\circ}\text{N}$ )位于格陵兰岛西南方向的洋面上;预报量与前一年 10 月的编号台风和当年 2 月的太阳黑子活动也有较好的相关性。

利用表 1 中的 6 个预测因子,采用 PLSR 算法,建立辽宁汛期平均降水量的预测模型。通过计算,1951 ~ 2001 年拟合和 2002 ~ 2011 年试预报结果的均方根误差分别为 87.26 mm 和 87.93 mm。可见,只使用环流量作为预报因子,模型拟合和试预报结果均不够理想,因此有必要对模型进行改进。

表 1 辽宁汛期平均降水量 PLSR 预报模型的预测因子

Tab. 1 The predictors of partial least square regression (PLSR) model for precipitation forecast during the flood season in Liaoning Province

| 预测因子                                               | 月份  | 相关系数  |
|----------------------------------------------------|-----|-------|
| 大西洋欧洲环流型                                           | -3  | 0.492 |
| 亚洲区极涡面积指数                                          | -12 | 0.439 |
| 北半球极涡面积指数                                          | -12 | 0.419 |
| SST( $54^{\circ}\text{W}$ 、 $66^{\circ}\text{N}$ ) | 2   | 0.407 |
| 编号台风                                               | -10 | 0.406 |
| 太阳黑子                                               | 2   | 0.404 |

注:负的月份表示前一年因子的月份,正值为当年的因子月份

#### 3.2 预报量的周期特征及其周期性因子

##### 3.2.1 预报量周期特征

利用 1951 ~ 2011 年辽宁汛期平均降水量时间序列,采用 Morlet 小波分析方法<sup>[16]</sup>,分析其周期变化的局部特征。由于时间序列相对较短,为了消除边界效应,采用周期延伸的方法对原始资料进行延拓,在原始数据的前后各延长一个周期,进行小波变换后再截取对应原始资料的小波系数来分析。

由图 2a 可见,1951 ~ 1958 年和 1978 ~ 1994 年小波系数实部正负间隔较明显,表明在这 2 个时期内辽宁汛期平均降雨量存在 4 ~ 6 a 的年际尺度准周期。此外,8 ~ 12 a 小波系数实部正负间隔明显,

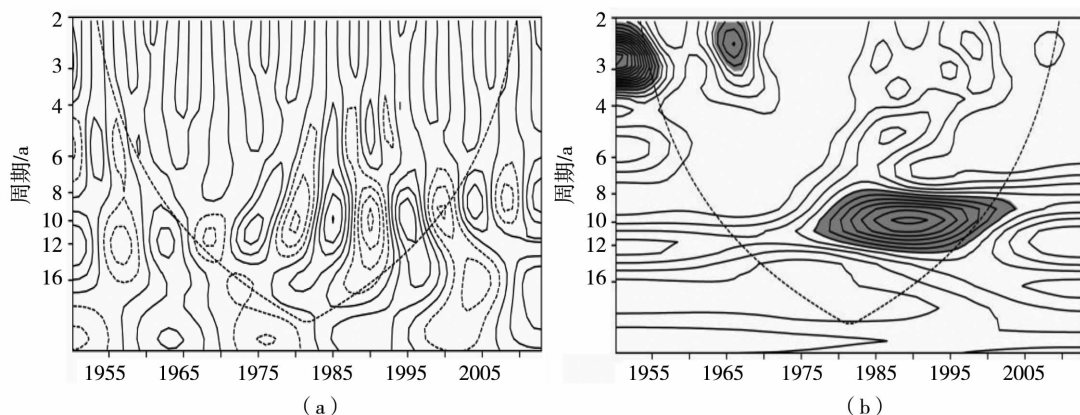


图 2 1951 ~ 2011 年辽宁汛期降雨量小波分析实部(a,实线为正值,虚线为负值)与波谱能量(b,阴影部分通过 95% 的信度检验)

(U 形虚线为  $10^{-5}$  误差等值线,其中内部误差  $< 10^{-5}$  表示受边界影响小;外部误差  $> 10^{-5}$  受边界影响较大)

Fig. 2 The real part (a, the solid and dashed lines for positive and negative values, respectively) and power spectrum (b, the shaded passed 0.05 confidence test) of Morlet wavelet analysis for precipitation during the flood season from 1951 to 2011 in Liaoning Province  
(The U - shape dashed curve for isolines with  $10^{-5}$  error, the errors inside the curve affected by the boundary were small, while those outside the curve were bigger)

其中 1953、1963、1975、1985、1995 和 2005 年的小波实部为正位相,1958、1968、1980、1990、2000 和 2008 年为负位相,说明辽宁汛期平均降雨量还存在 8 ~ 12 a 的长周期。由图 2b 可知,波谱能量主要集中在 2 ~ 3 a 和 8 ~ 12 a 波段上,其中 2 ~ 3 a 的波谱能量主要集中在 1951 ~ 1956 年和 1963 ~ 1968 年,且后者通过了 95% 的信度检验;8 ~ 12 a 的波谱能量主要集中在 1977 ~ 2004 年,且通过了 95% 的信度检验。这与王锦贵等<sup>[4]</sup>利用 20 世纪 90 年代以前资料得到的辽宁夏季降水存在 2 ~ 3 a、4 ~ 6 a 和 10 ~ 12 a 的周期振荡相同。

3.2.2 预报量周期性因子选用

利用 1951 ~ 2001 年建模样本集中的预报量时间序列,采用均生函数 (Mean Generating Function, MGF) 方法<sup>[17]</sup>筛选合适的周期因子。取最大周期  $m = [n/3]$  ( $[ ]$  表示取整,  $n$  为样本数), 这里  $n = 51$ , 因此构建了 17 个 MGF 延拓序列, 各序列与辽宁汛期平均降水量的相关性  $t$  检验分布如图 3 所示。从图 3 可知, 不同周期延拓 MGF 与预报量的相关性存在 3 个峰值, 分别为 5 a、10 a 和 16 a。其中, 5 a 序列的相关性  $t$  检验值为 2.06, 通过了  $\alpha = 0.05$  显著性检验; 10 a 序列的相关性  $t$  检验值为 3.806, 通过了  $\alpha = 0.01$  显著性检验; 16 a 序列的相关性  $t$  检验值最高, 为 5.725, 远高于  $\alpha = 0.01$  显著性检验阈值, 但由于周期长、样本数少, 随机性较大<sup>[16]</sup>。结合小波分析和 MGF 延拓序列  $t$  检验结果, 将周期为 5 a 和 10 a 的延拓均生函数序列作为 PLSR 预测模型的待选因子。

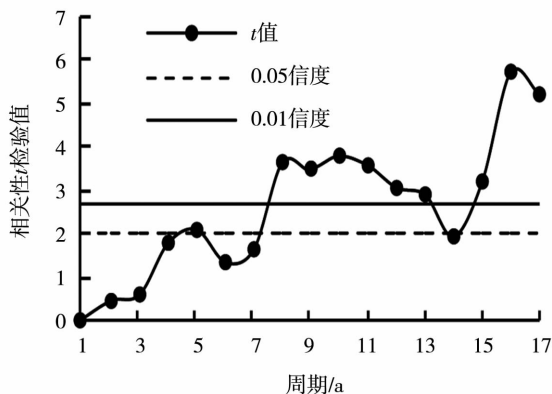


图 3 不同周期均生函数(MGF)序列与 1951 ~ 2001 辽宁汛期平均降雨量相关性  $t$  检验值分布

Fig.3 The distribution of  $t$ -test-value of correlation between the mean generating function with different periods and precipitation during the flood season in Liaoning Province from 1951 to 2001

3.3 预报量周期性因子对 PLSR 预测模型的改进

引入汛期平均降水量延拓均生函数因子进入 PLSR 预测模型, 得到的预测模型因子见表 2。可以看出, 在 5 a 和 10 a 均生函数序列中, 后者入选。与表 1 相比, 相关系数为 0.482 的均生函数序列(波数为 10)代替了相关系数为 0.404 的太阳黑子。改进后的预报模型 1951 ~ 2001 年拟合和 2002 ~ 2011 年试预报结果的均方根误差分别为 76.56 mm 和 77.85 mm, 较改进前的 PLSR 模型的误差下降约 10.0 mm, 改进效果较明显。

表 2 辽宁汛期平均降水量考虑均生函数的 PLSR 模型的预测因子

Tab.2 The predictors of PLSR model considering MGFs for precipitation forecast during the flood season in Liaoning Province

| 因子名称            | 月份(波数) | 相关系数  |
|-----------------|--------|-------|
| 大西洋欧洲环流型        | -3     | 0.492 |
| 均生函数序列          | 10*    | 0.482 |
| 亚洲区极涡面积指数       | -12    | 0.439 |
| 北半球极涡面积指数       | -12    | 0.419 |
| SST(54°W, 66°N) | 2      | 0.407 |
| 编号台风            | -10    | 0.406 |

注: 月份说明同表 1, \* 表示均生函数的波数

4 辽宁 5 站汛期降水量预测模型检验

辽宁省 5 站汛期降水量的周期与辽宁平均汛期降水量有一定程度的差异, 各站的预报量周期性因子分别从相应站的 4 ~ 6 a、8 ~ 12 a 波段的  $t$  检验峰值中选取, 构造均生函数延拓序列。利用 1952 ~ 2001 年 50 个样本, 引入环流指数、SST 和均生函数等因子, 分别建立沈阳、朝阳、营口、丹东和大连 5 站汛期降水量预测模型。然后, 利用 2002 ~ 2011 年独立样本, 对各站的预测效果进行  $P_s$  检验, 并与逐步回归模型的预报效果进行对比。其中, SWR 模型采用逐步回归方法建立 5 站汛期降水量预测方程, 其建模样本和检验样本的时间长度与 PLSR 模型一致, 待选因子中包含环流指数、SST 和均生函数等。

表 3 给出 PLSR 和 SWR 模型对辽宁地区 5 站 2002 ~ 2011 年汛期降雨量  $P_s$  评分。可以看出, PLSR 模型预测较好的年份有 6 a,  $P_s$  值  $\geq 80\%$ , 预测较差的年份有 4 a,  $P_s$  值  $\leq 60\%$ ; 而 SWR 模型预测较好的年份有 3 a,  $P_s$  值  $\geq 80\%$ , 预测较差的年份有 5 a,  $P_s$  值  $\leq 60\%$ 。此外从 10 a 平均来看, PLSR

和 SWR 模型的平均  $P_s$  评分分别为 72.6% 和 62.3%, 前者比后者高出 10.3%。整体而言, PLSR 模型对汛期降水量的预测效果明显好于 SWR 模型。

表3 PLSR 模型与 SWR 模型对 2002 ~ 2011 年  
辽宁 5 站汛期降雨量预测的  $P_s$  评分(单位:%)

Tab.3 The  $P_s$  scores of PLSR and SWR models considering  
MGF for precipitation forecast during the flood season in 5  
stations of Liaoning Province from 2002 to 2011 (Unit:%)

| 年份   | PLSR  | SWR   |
|------|-------|-------|
| 2002 | 100.0 | 80.0  |
| 2003 | 81.4  | 40.0  |
| 2004 | 60.0  | 71.4  |
| 2005 | 100.0 | 60.0  |
| 2006 | 20.0  | 00.0  |
| 2007 | 60.0  | 60.0  |
| 2008 | 80.0  | 80.0  |
| 2009 | 60.0  | 71.4  |
| 2010 | 80.0  | 100.0 |
| 2011 | 85.7  | 60.0  |
| 平均   | 72.6  | 62.3  |

## 5 结 论

(1) 利用 1951 ~ 2011 年全国 160 站逐月降水资料, 对辽宁省汛期平均降水量气候特征进行了初步分析。得到在 61 a 间汛期降水呈逐年下降趋势, 有明显的年代际变化特征; 结合小波分析结果显示, 辽宁汛期平均降水量存在 2 ~ 3 a、4 ~ 6 a 和 8 ~ 12 a 的准周期。这些结论, 为将周期性因子引入辽宁省汛期降水量预测模型提供了理论依据。

(2) PLSR 方法是一种新型回归技术, 能够有效克服预报因子之间的多重相关性影响。对辽宁 5 站汛期降水量预测试验结果表明, 在 10 a 试报中, PLSR 模型  $P_s$  评分平均值达 72.6%, 比 SWR 模型提高了 10.3%, 表明 PLSR 方法较普通回归方法有明显优势, 具有很好的应用推广前景。

(3) 采用 MGF 方法, 将辽宁汛期平均降水量 5 a 和 10 a 周期变化, 延拓成 MGF 序列, 作为辽宁汛

期平均降水量 PLSR 预测模型的待选因子, 可有效提升模型的预测效果。试验中, 引入了均生函数因子的 PLSR 模型, 比没有引入的模型均方根误差降低了约 10.0 mm。说明 MGF 构造预报量的周期性因子, 在一定程度上弥补了气候预测统计模型高相关性因子不足的问题, 使模型具有更好的物理基础。

## 参考文献:

- [1] 丁一汇, 戴晓苏. 中国近百年来的温度变化[J]. 气象, 1994, 20(12): 19 - 26.
- [2] 孙莹, 万丽岩, 江静. 辽宁降水分区变化特征及夏季降水影响因素分析[J]. 气象与环境学报, 2008, 24(3): 18 - 23.
- [3] 杨文艳, 迟春艳. 辽宁省四季降水时空变化特征分析[J]. 安徽农业科学, 2008, 36(3): 9197 - 9199.
- [4] 王锦贵, 许君强. 东北短期气候特征研究[M]. 北京: 气象出版社, 1998. 7 - 11.
- [5] 李辑, 王景贵. 东北夏季低温与旱涝预测综合业务系统[J]. 气象科技, 2002(3): 252 - 256.
- [6] 夏梅艳, 赵连伟. 辽宁 1999 年夏季少雨干旱气候成因分析[J]. 辽宁气象, 2000(3): 11 - 14.
- [7] 杨文艳. 辽宁汛期降水模糊聚类分析[J]. 辽宁气象, 2004(1): 7 - 8.
- [8] 李广霞, 陈传雷, 才奎志. 辽宁夏季降水变化特征分析[J]. 气象与环境科学, 2008, 31(2): 31 - 34.
- [9] 王慧文. 偏最小二乘回归方法及其应用[M]. 北京: 国防工业出版社, 1999. 274.
- [10] Wold S, Albano C, Dunn M, et al. Pattern regression finding and using regularities in multivariate data[C]. In Martens J. In Proc. IUFOST Conf. "Food Research and Data", London, Analysis Applied Science Publication, 1983.
- [11] Wold S, Trygg J, Berglund A, et al. Some recent developments in PLS modeling[J]. Chemometrics and Intelligent Laboratory Systems, 2001, 58: 131 - 150.
- [12] Hoskuldson A. PLS regress methods[J]. Journal of Chemometrics, 1988, 2: 211 - 228.
- [13] Geladi P, Qkowlaski B. Partial Least Squares regression: A tutorial[J]. Analytica Chemical Acta, 1986, 35: 1 - 17.
- [14] 黄颖, 金龙. 华南前汛期降水预测模型及其预测试验[J]. 热带气象学报, 2011, 27(5): 753 - 757.
- [15] 金龙, 罗莹, 王业宏, 等. 月降水量的神经网络混合预报模型研究[J]. 高原气象, 2003, 22(6): 618 - 623.
- [16] 张冰, 巩远发, 徐影, 等. CMIP5 全球气候模式对中国地区干旱变化模拟能力评估[J]. 干旱气象, 2014, 32(5): 694 - 700.
- [17] 魏凤英. 现代气候统计诊断与预测技术[M]. 北京: 气象出版社, 2007. 296.

## Application of Partial Least Square Regression on Precipitation Prediction During the Flood Season in Liaoning Province

ZHAO Zhongjun<sup>1</sup>, LIU Shanliang<sup>2</sup>, YOU Daming<sup>3</sup>,

WANG Xuezhong<sup>3</sup>, HU Banghui<sup>3</sup>, CHENG Yifan<sup>4</sup>

(1. Troop Unit 92493 of PLA, Huludao 125000, China; 2. Troop Unit 61741 of PLA, Beijing 100081, China;  
3. Institute of Meteorology and Oceanography, PLAUST, Nanjing 211101, China;  
4. College of Atmospheric Science, Lanzhou University, Lanzhou 730000, China)

**Abstract:** Based on the monthly mean temperature and precipitation at 160 weather stations in China during 1951–2011, 74 circulation indexes from National Climate Center and the monthly SST from NCEP/NCAR, combined with mean generating function (MGF), the prediction models of precipitation during the flood season in 5 stations of Liaoning Province were respectively established by using the partial least square regression (PLSR). And the effects predicted by PLS model on precipitation were tested. The result showed that the periodic factor of the prediction founded by MGF could weaken the correlation of the predictors in the statistic model to some extent. The root mean square error (RMSE) of mean precipitation during the flood season in Liaoning Province predicted by PLSR model considering the periodicity from 2002 to 2011 was reduced by 10.0 mm. The effect predicted by PLSR model on precipitation during the flood season was much more efficient in comparison with that predicted by step wise regression (SWR) model owing to better resolve about the multi-correlation problem. The mean prediction skill score of precipitation during the flood season in 5 stations from 2002 to 2011 was 72.6%, which improved by 10.3% than that of SWR model.

**Key words:** partial least square regression; mean generating function; precipitation prediction during the flood season

(上接第 999 页)

## Temporal and Spatial Variation of Shallow Groundwater Level and Its Response to Precipitation in the Mid-west of Jilin Province

MA Yanmin, LI Jianping, WANG Ying, CHEN Liwen, TANG Xiaoling, LIU Ke

(Institute of Meteorological Sciences of Jilin Province, Changchun 130062, China)

**Abstract:** Groundwater played an important role in the agriculture production and people livelihood, and it was necessary to study the dynamic changes of groundwater. Based on the monthly precipitation and shallow groundwater level of 19 observation stations during 2007–2013, the temporal and spatial distribution characteristics of shallow groundwater level and their response to precipitation in the mid-west of Jilin Province were comparatively analyzed. Results are as follows: (1) The annual average shallow groundwater level in mid-western Jilin during 2007–2013 increased gradually from northeast to southwest as a whole, and the dynamic changes of groundwater level in the middle and west of Jilin were distinctly different. The monthly fluctuation of groundwater level in most stations of the west region wasn't obvious, but the annual change trends from 2007 to 2013 were significantly different. However, the monthly changes in stations of the middle region were relatively bigger, and the annual average groundwater levels in other stations during 2007–2013 increased except for Changchun station. (2) The correlations between the annual and monthly average shallow groundwater level and precipitation in each station of the central region were positive, thus the recharge of shallow groundwater was mainly derived from the rainfall infiltration. However, the correlations in stations of the west region were almost negative or weren't significant, and the effects of artificial exploitation on shallow groundwater level were larger. (3) The effects of precipitation on groundwater level presented in three respects: the change characteristics of annual shallow groundwater level were similar to annual precipitation, and precipitation affected the range of shallow groundwater level. The change trend and range of underground water level before and after flood season were closely related to precipitation in flood season. The precipitation during flood season was helpful to offset groundwater exploitation in spring, but the mitigation was slightly different in wet and dry years.

**Key words:** shallow groundwater level; temporal and spatial variation; precipitation; correlation; the mid-west of Jilin Province