

基于气候重建资料及均生函数-最优子集回归模型的降水预测

马 龙, 刘廷玺, 冀鸿兰, 高瑞忠

(内蒙古农业大学水利与土木建筑工程学院, 内蒙古 呼和浩特 010018)

摘 要: 利用由科尔沁沙地榆树区域综合年表而重建的 181a 降水序列, 采用均生函数-最优子集回归 (MGF-OSR) 预测模型对科尔沁沙地降水进行了 5a 情景的预测, 克服了以往序列短的不足。模型建模期相对误差绝对值在 3%、5%、10% 以内的年数分别占建模年数的 36.46%、64.64%、93.37%, 建模期模拟精度较高。2007~2010 年实测检验期, 模拟与实测值保持了方向一致性, 各年份相对误差绝对值分别为 11.40%、6.73%、22.06%、5.49%, 大部分年份预测较理想。经预测, 2011~2015 年期间, 2011 年、2012 年、2014 年和 2015 年为降水正常年份, 2013 年为降水偏干年份。与实际情况对比, 2011 年与预测一致, 2012 年稍有差别。总体上, 2011~2015 年降水变化较为平缓。

关键词: 气候重建资料; 均生函数; 最优子集回归; 科尔沁; 降水; 预测

中图分类号: P333

文献标识码: A

文章编号: 1000-0852(2013)01-0063-05

1 引言

科尔沁沙地在历史上曾为水草丰美、植被茂盛的大草原。经过近百年的草场退化、沙化及土地荒漠化, 它已成为中国四大沙地之一, 其地表环境的变化不论在全球还是中国, 都具有典型性和代表性。这其中水文气候要素变化对研究区生态环境变化起到了重要作用。研究水文要素变化规律至关重要。在科尔沁沙地及其周边地区相关研究中^[1-3], 研究时段大部分为 20 世纪 50 年代至 20 世纪末。研究内容涉及降水、平均气温、沙尘暴发生日数、平均最高和最低气温、极端最高和最低气温及降水极端事件等要素的变化特征及变化趋势等。

然而, 这些研究大部分以气象观测站网观测数据为基础, 资料年限一般为 50a 左右。资料年限短, 站点使用少, 不能涵盖整个区域, 对研究区域更长时间尺度的水文气候变化显然是不够的, 而树木年轮记录揭示的水文气候信息是解决这一问题的有效手段之一。树木年轮具有定年准确、连续性好、分辨率高及与气候变

化联系紧密且重建精度高的特点^[4], 为过去水文资料的获取提供了可靠依据, 其中马龙等^[5]以科尔沁沙地榆树为研究对象, 首次分析了科尔沁沙地榆树生长与水文气象因子的关系, 并对水文气候进行了重建。

大气降水是一种复杂的自然现象, 进行预报是目前一项研究难题。国内外在该方面作了大量研究, 目前常用的预测方法有概率统计法、均生函数法、小波变换法、人工神经网络法、时间序列方法、Markov 法、光滑支持向量机等方法^[6-7], 大部分方法主要存在的是精度问题, 而均生函数法可以改善其它模型对序列极值预测的不足, 提高极值拟合与预测效果。

基于上述情况, 本文以科尔沁沙地区域榆树年表重建的 1826~2006 年共计 181a 降水序列为基础, 采用均生函数-最优子集回归 (MGF-OSR) 预测模型, 预测科尔沁沙地 2011~2015 年 5a 的降水变化。

2 资料与方法

2.1 研究区概况

科尔沁沙地年降水量在 200~650mm 之间, 70% 的

收稿日期: 2011-10-16

基金项目: 国家自然科学基金项目 (50869005); 内蒙古自然科学基金项目 (2010BS0608); 内蒙古农业大学“水资源利用”创新团队计划

作者简介: 马龙 (1978-), 男, 内蒙古呼和浩特人, 副教授, 博士, 硕士生导师, 主要从事水文及水资源、环境演变、气候变化及与环境响应关系研究。

E-mail: malong4444333@163.com

降水集中在6、7、8三月份;年蒸发量在1 600~2 400mm(蒸发皿为20cm口径)之间,主要集中在4~9月,此间蒸发量占全年蒸发量的78%左右;夏季炎热,冬季漫长干冷,多年平均气温为6℃左右。现状地貌形态主要包括固定沙地、半固定沙地和流动沙丘。风沙土是科尔沁沙地面积最大的土类,其次是草甸土,再次是栗钙土,此外还有盐碱土、暗棕壤土、黑钙土等。植被区系处于蒙古、长白和华北植物区系的交汇地,植物种类较为丰富。

2.2 使用的降水资料

本文使用的降水资料为由本课题组最近在科尔沁沙地建立的区域榆树树轮宽度标准年表而重建的降水序列(1826~2006年,共计181a),重建的降水序列见图1^[5]。

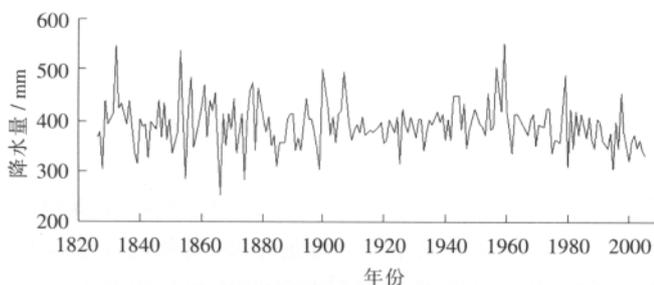


图1 区域1826~2006年年降水量重建序列分布

Fig.1 Reconstruction series distribution of annual precipitation from 1826 to 2006

上述降水重建过程中使用的实测降水序列是采用10个采样点各自距离最近的气象站点进行平均,即8个气象站点的平均值序列(8个气象站点为科左后旗、通辽、开鲁、科左中旗、科右中旗、扎鲁特旗、巴林右旗、库伦旗)。因此,也同样收集了这8个气象站点自建站至2010年月降水序列,以同样计算方式计算出区域降水量,以作实测与模拟对比分析之用。

2.3 均生函数-最优子集回归(MGF-OSR)预测模型

均生函数(Mean Generating Function, MGF)是时间序列均值生成函数的简称,目前被广泛应用于回归分析中,同时给出相应的建模方案,使回归模拟精度较高。均生函数-最优子集回归(MGF-OSR)预测模型过程如下^[8]:

(1)计算时间序列 $x(t) = \{x(1), x(2), \dots, x(n)\}$ (n 为样本的总量)的均值:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x(i) \quad (1)$$

对于均值,定义其生成函数为:

$$\bar{x}_l(i) = \frac{1}{n} \sum_{j=0}^{n-1} x(i+jl) \quad (2)$$

式中: $i=1, \dots, l; 1 \leq l \leq m; n_l = INT(n/l); m = INT(n/2)$ 或 $m = INT(n/3); INT$ 表示取整数。鉴于短周期均生函数随机性小的特点,本次取 $m = INT(n/3)$, 由此,可计算得到 $x(t)$ 序列的均生函数 $\bar{x}_l^{(0)}(t)$ 。

(2)为拟合原序列中的高频部分,对于原始序列 $x(t)$ 进行一阶及二阶差分运算,得到:

$$\Delta x(t) = x(t+1) - x(t), t=1, 2, \dots, n-1$$

$$\Delta^2 x(t) = \Delta x(t+1) - \Delta x(t), t=1, 2, \dots, n-2$$

同样,根据式(2)计算一阶及二阶差分序列的均生函数,即为 $\bar{x}_l^{(1)}(t)$ 和 $\bar{x}_l^{(2)}(t)$ 。

(3)根据方程(2),每一个序列可以得到 m 个均生函数,形成矩阵 H 。

$$H = \begin{pmatrix} \bar{x} \\ \bar{x}_2(1), \bar{x}_2(2) \\ \bar{x}_3(1), \bar{x}_3(2), \bar{x}_3(3) \\ \cdot \\ \cdot \\ \cdot \\ \bar{x}_m(1), \bar{x}_m(2), \dots, \bar{x}_m(m) \end{pmatrix} \quad (3)$$

由矩阵 H 可知,均生函数即由时间序列按一定的时间间隔计算均值而派生出的。由均生函数矩阵 H ,将其定义域延拓到整个数轴上,也就是做周期性延拓:

$$f_i(t) = \bar{x}_l(i) \quad (4)$$

式中: $t \equiv i \pmod{l}$, mod 表示同余; $t=1, 2, \dots, n$ 。

称 $f_i(t)$ 为均生函数延拓序列,由此可构造出均生函数延拓矩阵 F :

$$F = (f_{ij})_{n \times mn}, f_{ij} = f_i(t) \quad (5)$$

$$F = \begin{pmatrix} \bar{x}, \bar{x}_2(1), \bar{x}_3(1), \dots, \bar{x}_m(1) \\ \bar{x}, \bar{x}_2(2), \bar{x}_3(2), \dots, \bar{x}_m(2) \\ \bar{x}, \bar{x}_2(1), \bar{x}_3(3), \dots, \cdot \\ \bar{x}, \bar{x}_2(2), \bar{x}_3(1), \dots, \cdot \\ \cdot \\ \cdot \\ \cdot \\ \bar{x}, \bar{x}_2(i_2), \bar{x}_3(i_3), \dots, \bar{x}_m(i_m) \end{pmatrix}$$

式中: $\bar{x}_2(i_2)$ 表示顺序取 $\bar{x}_2(1)$ 、 $\bar{x}_2(2)$ 之一, $\bar{x}_3(i_3)$ 表示顺序取 $\bar{x}_3(1)$ 、 $\bar{x}_3(2)$ 、 $\bar{x}_3(3)$ 之一, 余下类推。这里称 f_i 为延拓均生函数。

由此, 分别得到原序列、原序列的一阶及二阶差分序列的延拓序列 $f_i^{(0)}(t)$ 、 $f_i^{(1)}(t)$ 和 $f_i^{(2)}(t)$ 。

(4) 为了拟合时间序列中向上递增和向下递减的趋势, 建立累加延拓序列:

$$f_i^{(3)}(t) = x(1) + \sum_{i=1}^{t-1} f_i^{(1)}(i+1) \quad (6)$$

式中: $l=1, 2, \dots, m$ 。当 $t=1$ 时, $f_i^{(3)}(t)=x(1)$ 。

(5) 由前述过程共获得 $4m$ 个均生函数延拓序列, 即 $f_i^{(0)}(t)$ 、 $f_i^{(1)}(t)$ 、 $f_i^{(2)}(t)$ 和 $f_i^{(3)}(t)$ ($l=1, 2, \dots, m$), 将其作为自变量供筛选。

计算双评分准则:

$$CSC = S_1 + S_2 \quad (7)$$

式中: S_1 为数量评分, $S_1 = nR^2$; S_2 为趋势评分: $S_2 = 2I = 2 \left[\sum_{i=1}^I \sum_{j=1}^I n_{ij} \ln n_{ij} + n \ln n - \left(\sum_{i=1}^I n_i \ln n_i + \sum_{j=1}^I n_j \ln n_j \right) \right]$ 式中 I 为预报趋势类别项, n_{ij} 为 i 类事件与第 j 类估计事件列联表的个数。

(6) 初步粗选预测因子。将所有延拓序列与原序列进行一元回归, 计算 CSC 值, 凡 $CSC > \chi^2$ 的序列粗选为预报因子, 设入选了 P 个延拓序列。

(7) 计算所有可能的 2^P 个回归子集的 CSC 值, 选取 CSC 值最大的为最优回归子集, 由此, 建立预测模型, 作出步长为 5 年的预测。

3 结果及分析

降水重建序列年限为 1826~2006 年, 共计 181 年, 按照前述过程(1)~(5), 得到 4 个均生函数延拓序列, 根据过程(5)~(7), 计算出不同变量个数情况下的最优子集, 选取值最大的为最优回归子集, 由此, 建立预测模型, 作出步长为 5a 的预测。计算获得预测方程如下:

$$\begin{aligned} x_{yc}(t) = & -23379.57 + 0.32f_{52}^{(0)}(t) + 0.28f_{53}^{(0)}(t) + 0.35f_{54}^{(0)}(t) \\ & + 1441.51f_{57}^{(0)}(t) + 0.43f_{59}^{(0)}(t) + 0.34f_{60}^{(0)}(t) - 0.19f_{50}^{(1)}(t) + \\ & 1441.36f_{57}^{(1)}(t) - 0.03f_{52}^{(2)}(t) - 0.02f_{57}^{(2)}(t) + 0.01f_{19}^{(3)}(t) + \\ & 0.09f_{25}^{(3)}(t) - 1441.10f_{57}^{(3)}(t) \end{aligned}$$

这里将 1826~2006 年作为降水重建序列建模阶段, 而 2007~2010 年是模拟与实测拟合检验阶段, 2011~2015 年为预测阶段。整个时段为模拟阶段。由上述模型模拟后, 在建模阶段, 模拟值与原重建序列相关系数, 模拟结果与原重建序列、实测序列拟合情况及未来预测见图 2。

由图 2 可以看出, 建模阶段降水重建值与模拟值

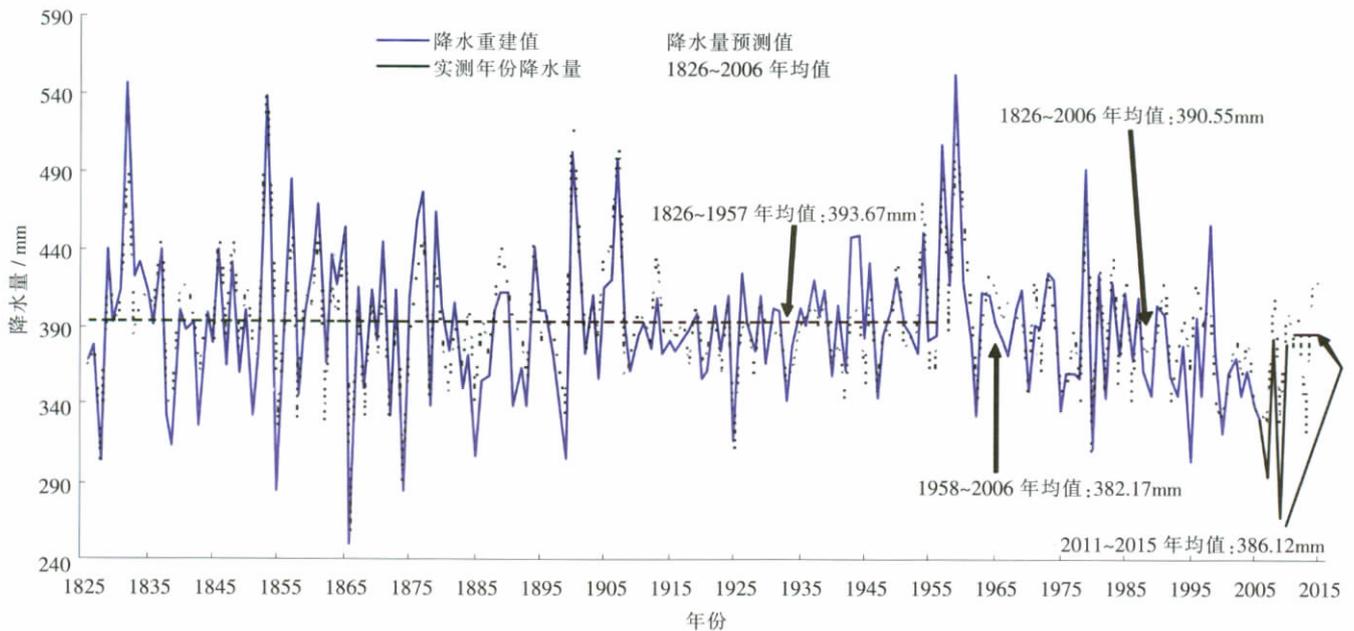


图 2 降水重建序列、实测序列与模拟序列对比及未来降水预测图

Fig. 2 Comparison between reconstructed series, observed series and simulated series and the future precipitation forecasting

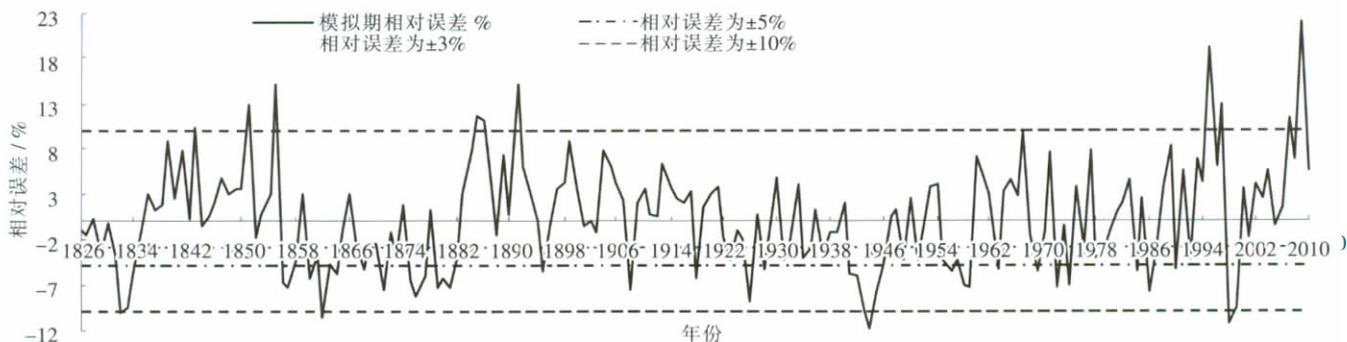


图3 模拟期相对误差分布

Fig.3 Distribution of the relative errors in the simulation period

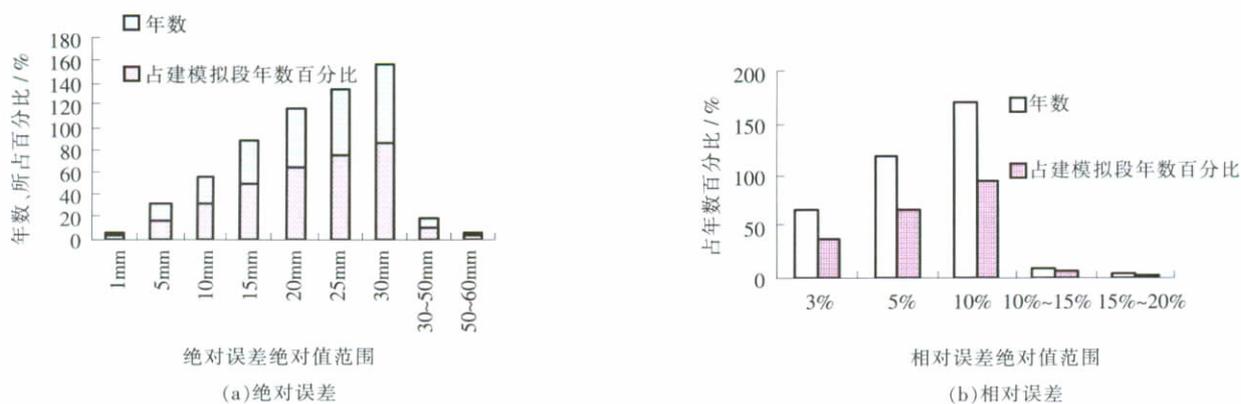


图4 建模期相对误差、绝对误差绝对值分布

Fig.4 Absolute value distribution of the relative error and absolute error in the modeling period

吻合较好,特别是峰值拟合较为理想。模拟期相对误差分布见图3,建模期相对误差、绝对误差绝对值分布见图4。由图3和图4可知,建模期相对误差绝对值均值为4.49%,最大值为19.23%(1995年),最小值为0.09%(1902年);相对误差绝对值在3%、5%、10%以内的年数分别为66a、117a、169a,分别占建模年数的36.46%、64.64%、93.37%;相对误差绝对值在10%~15%的年数有9a,占建模年数的5%;相对误差绝对值在15%~20%的年数有3a,占建模年数的1.7%。在建模期绝对误差中,绝对误差绝对值均值为17.42mm,最大值为58.36mm,最小值为0.32mm;绝对误差绝对值在1mm、5mm、10mm、15mm、20mm、25mm、30mm以内的年份分别为6a、31a、57a、90a、116a、135a、157a,分别占建模年数的6.31%、17.13%、31.49%、49.72%、64.09%、74.59%、86.74%;在30~50mm、50~60mm之内的年份分别为19a、5a,分别占建模年数的10.50%、2.76%。总体上,模型建模期模拟精度较高,符合水文情报预报规范对预报的精度评定要求。所建立的模型可以用于预测。

2007~2010年检验期模拟值与实测值对比见图

2,相对误差分布见图3。由图2、3可知,模拟值与实测值变化保持了较好的方向一致性。2007~2010年绝对误差绝对值分别为33.56mm、25.80mm、59.28mm、20.85mm,均值为34.87mm;相对误差绝对值分别为11.40%、6.73%、22.06%、5.49%,均值为11.42%。这里,除2009年预测效果不太理想外,其余预测均较好。总体上,预测结果可用。

利用上述所建模型对2011~2015年5a降水的预测结果见图2。以整个时段(1826~2015年)序列均值为基础,采用降水距平百分率进行干湿变化频率分析,这里,以降水距平百分率在-10%~10%的年份为正常年份,以>30%的年份为湿润年,以<-30%的年份为干旱年,其他期间年份分别为偏湿年和偏干年。经计算,2011~2015年降水距平百分率分别为-2.57%、1.54%、-16.97%、6.58%、7.71%,其中2011年、2012年、2014年和2015年为降水正常年份,只有2013年为降水偏干年份。2011年和2012年实际降水准确数据虽未公布,但比对实际情况,2011年为降水正常年份,与预测一致;2012年为偏湿年,与预测稍有差别。

总体上,2011~2015年降水变化较为平缓,没有显著的趋势。

4 结论与讨论

(1) 该方法的实质是通过原序列与其本身隐含的各种周期变化的关系来构建预测模型,这就要求原序列具有足够的长度,长度越长,代表性越强。而本次使用1826~2006年181a降水重建资料,克服了以往序列短的不足,符合这点要求,使建模及预测更加趋于实际。

(2) 建模期相对误差绝对值在3%、5%、10%以内的年数分别占建模年数的36.46%、64.64%、93.37%,在10%~20%的年数只占建模年数的6.7%。总体上,模型建模期模拟精度较高,模型可以用于预测。

(3) 对于2007~2010年检验期,模拟值与实测值变化方向一致,大部分年份预测均较理想。

(4) 2011~2015年期间,2011年、2012年、2014年和2015年为降水正常年份,只有2013年为降水偏干年份。与实际情况对比,2011年与预测一致,2012年稍有差别。总体上,2011~2015年降水变化较为平缓。

(5) 预测模型在考虑原序列周期性变化的基础上,对其一二阶差分序列也作了考虑,较为理想的进行了预测。但没有考虑序列变化的影响因子,今后应加强结合机理性过程。

参考文献:

[1] 朱志伟, 花振飞. 华北地区冬季降水时空分布特征分析 [J]. 水文,

2010, 30 (4): 84-87. (ZHU Zhiwei, HUA Zhenfei. Characteristics analysis of winter precipitation in North China [J]. Journal of China Hydrology, 2010, 30(4): 84-87. (in Chinese))

[2] 王红梅, 骆永明, 蒋德明. 科尔沁沙地翁牛特旗气温的非对称变化分析[J]. 干旱区研究, 2007, 24(2): 261-267. (WANG Hongmei, LUO Yongming, JIANG Deming. Unsymmetrical variation of the maximum and minimum air temperatures in Ongniud Banner, Horqin sandland [J]. Arid Zone Research, 2007, 24(2): 261-267. (in Chinese))

[3] 渠翠平, 关德新, 王安志, 等. 近56年来科尔沁沙地气候变化特征[J]. 生态学杂志, 2009, 28 (11): 2326-2332. (QU Cuiping, GUAN Dexin, WANG Anzhi, et al. Characteristics of climate change in Horqin sandy land in past 56 years. [J]. Chinese Journal of Ecology, 2009, 28(11): 2326-2332. (in Chinese))

[4] Fritts H C. Tree Ring and Climate [M]. London: Academic Press, 1976, 1-500.

[5] 马龙, 刘廷玺, 冀鸿兰, 等. 科尔沁沙地1826年以来榆树树轮宽度年表的建立及降水重建[J]. 第四纪研究, 2011, 31(2): 360-369. (MA Long, LIU Tingxi, JI Honglan, et al. The establishment of *Ulmus pumila* L. ring width chronology and rainfall reconstruction in Horqin sandy land since 1826 [J]. Quaternary Sciences, 2011, 31 (2): 360-369. (in Chinese))

[6] Widmann M, Bretherton C S, Salathé E P. Statistical precipitation downscaling over the northwestern United States using numerically simulated precipitation as a predictor [J]. Journal of Climate, 2003, 16(5): 799-816.

[7] Wetterhall F, Halldin S, Xu C. Statistical precipitation downscaling in central Sweden with the analogue method [J]. Journal of Hydrology, 2005, 306(1-4): 174-190.

[8] 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: 239-259. (WEI Fengying. Modern Climatic Statistical Diagnosis and Prediction Technology [M]. Beijing: China Meteorological Press, 2007: 239-259. (in Chinese))

Precipitation Short-term Forecasting Based on Climate Reconstruction Data and Mean Generating Function-optimal Subset Regression (MGF-OSR) Prediction Model

MA Long, LIU Tingxi, JI Honglan, GAO Ruizhong

(College of Water Conservancy and Civil Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China)

Abstract: The future 5-year precipitation forecasting was made for Horchin sandy land by using Mean Generating Function-Optimal Subset Regression (MGF-OSR) based on the reconstructed series of 181 years. The years that the absolute values of relative errors less than of 3%, 5%, and 10% respectively account for 36.46%, 64.64%, 93.37% of the modeling years, and the simulation accuracy is higher in modeling period. The simulated values are consistent with observed data in the experiment period from 2007 to 2010, during which the absolute values of the relative errors are 11.40%, 6.73%, 22.06% and 5.49%, respectively. By forecasting, 2011, 2012, 2014 and 2015 are normal years for precipitation, and 2013 is a dry year for precipitation.

Key words: climate reconstruction data; mean generating function; optimal subset regression; Horqin; precipitation; forecast