

基于加权马尔可夫模型的条件植被温度指数预测研究

韩萍¹, 王鹏新^{2*}, 王家慧¹, 孙威², 朱德海²

(1. 中国农业大学理学院, 北京 100094; 2. 中国农业大学信息与电气工程学院, 北京 100083)

摘要: 条件植被温度指数(VTCI)是一种近实时的干旱监测指标。利用加权马尔可夫模型对陕西关中平原的条件植被温度指数(VTCI)时空变化序列进行了 2 步预测, 并将预测结果与实际监测结果进行对比分析, 结果表明, 预测的干旱监测结果与实际的干旱监测结果空间分布特征基本一致, 进一步的误差分析显示, 2 步预测的误差平均值分别为 0.042 和 0.152。

关键词: 干旱预测; 条件植被温度指数; 加权马尔可夫模型

中图分类号: S161.2⁺¹ **文献标识码:** A **文章编号:** 1000-7601(2008)06-0196-05

干旱已给许多国家的粮食生产、水资源和能源造成了极大的影响。干旱在我国普遍存在, 年均受灾面积约 140 多万 hm^2 ^[1], 干旱造成的损失在各种自然灾害造成的损失中列为首位。因此对干旱进行客观、动态、实时监测和预测, 对于及时了解旱情程度和分布, 采取有效的防旱、抗旱措施具有重要的现实意义。

随着遥感技术的发展, 遥感提供的多光谱信息使人们可以利用植被指数和土地表面温度进行面上的干旱监测。植被指数可以反演植被的绿度和植物的生长状况, 进而用于监测植物的水分需求状况。其中应用最广的是归一化植被指数(NDVI), 在此基础上又出现了其它一些基于植被指数的干旱监测方法: 距平植被指数(AVI)、条件植被指数(VCI)等^[2]。土地表面温度(LST)可以反映土壤水分状况, 也常用于监测干旱。王鹏新等基于植被指数和土地表面温度的散点图呈三角形区域分布的条件, 提出了条件植被温度指数(VTCI)的干旱监测方法, 并将此方法用于陕西关中平原, 证实了 VTCI 与最近 2 旬的降水量密切相关^[3~5]。Wan 等将 VTCI 指数应用于美国南部大平原的德克萨斯州和俄克拉荷马州的干旱监测, 结果显示 VTCI 和该地区月累计降水量之间存在线性关系, 说明了 VTCI 是一种实时的干旱监测指标^[6]。

马尔可夫模型是时间序列预测中常用的方法, 它将时间序列看成一个随机过程, 通过对事物不同状态的初始概率和状态之间转移概率的研究, 确定状态变化趋势, 预测事物的未来。近年来马尔可夫模型已广泛应用于自然灾害过程的模拟和预测,

Paulo 等利用葡萄牙南部 Alentejo 7 个气象站 68 年的 SPI 数据, 运用马尔可夫模型和对数线性模型预报干旱发生的时间、级别以及周期^[7]。Wilks 运用马尔可夫模型对美国 30 个气象站的降水资料进行建模分析, 模拟年降水量和干湿周期的变化^[8]。孙才志等也以降水量为数据采用加权马尔可夫模型对降水丰枯状况进行了预测^[9]。本文利用加权马尔可夫模型对遥感面上获得的陕西关中平原 VTCI 干旱监测结果进行分析, 建立了预测模型并预测了 VTCI 的时空变化, 并对加权马尔可夫模型的预测结果进行了分析评估。

1 材料和方法

1.1 数据源

旱灾是陕西省的主要自然灾害之一, 陕西省中部的关中地区又是旱灾频发的地区。关中平原西起宝鸡市, 东至潼关, 南接秦岭, 北抵陕北高原, 东西长 300 km, 西狭东宽, 平均海拔 400 m 左右, 多年平均降雨量 500~700 mm^[10, 11]。

本研究应用关中平原 1999~2003 年 3~6 月的 AVHRR 多时段卫星遥感反演的 NDVI 和 LST 产品, 根据下面的方法计算 1999~2003 年 3~6 月各旬的 VTCI 监测数据, 得到关中平原的 VTCI 监测数据序列。

条件植被温度指数 (Vegetation Temperature Condition Index, VTCI) 的定义为^[3~5]:

$$VTCI = \frac{LST_{NDVI_i, \max} - LST_{NDVI_i}}{LST_{NDVI_i, \max} - LST_{NDVI_i, \min}} \quad (1)$$

收稿日期: 2008-02-20

基金项目: 国家自然科学基金项目(40571111); 教育部科学技术研究重点项目(105013)

作者简介: 韩萍(1968—), 女, 河南民权县人, 讲师, 在读博士, 主要从事干旱监测和预测等研究。E-mail: hanrose@cau.edu.cn.

* 通讯作者: 王鹏新(1965—), 男, 陕西礼泉县人, 副教授, 博士生导师, 主要从事定量遥感及其在农业中的应用研究。

其中, $LST_{NDVI_i, \max} = a + b \cdot NDVI_i$; $LST_{NDVI_i, \min} = a' + b' \cdot NDVI_i$ 。式中, $LST_{NDVI_i, \max}$ 和 $LST_{NDVI_i, \min}$ 分别表示在研究区域内, 当 $NDVI_i$ 值等于某一特定值时的土地表面温度的最大值和最小值; a 、 b 、 a' 、 b' 为四个待定系数, 可通过绘制研究区域的 NDVI 和 LST 的散点图近似地获得。四个待定系数的确定方法非常重要, 这里采用最大-最小值合成方法, 即应用多年的旬 NDVI 最大值合成产品和基于每年的旬 LST 最大值合成产品再合成的多年 LST 最小值产品确定冷边界 LST_{\min} 的系数 a' 和 b' ; 使用最大值合成技术, 即应用多年的旬 NDVI 和 LST 最大值合成产品确定热边界 LST_{\max} 的系数 a 和 b ^[5]。

1.2 加权马尔可夫模型

若系统的随机过程 X_t 在时刻 t 的状态用 E 表示, 那么在时刻 τ ($\tau > t$) 时, 系统所处状态与时间 t 以前所处状态无关, 则这一过程称为马尔可夫过程。马尔可夫模型将时间序列看作是时间离散、状态离散的马尔可夫过程, 即马尔可夫链。一系列相依的随机变量, 其各阶自相关系数刻画了各种滞时状态间相关关系的强弱, 若分别依其以前若干时段变量值的状态预测某时段的变量值的状态, 可按前面各时段与该时段相依关系的强弱进行加权求和, 然后再根据各状态加权求和的结果来预测该时段可能所处的状态。这就是加权马尔可夫模型预测的基本思想, 该方法充分、合理地利用了已有的信息。

在本研究中将 VTCI 时空序列看作具有不同空间分布的时间序列, 对每一个空间点的时间序列采用加权马尔可夫模型进行预测, 预测步骤如下^[7, 12, 13]:

(1) 确定马尔可夫链的状态空间 ($E = \{1, 2, \dots, N\}$)。

状态空间数目的确定直接影响最后预测值的精度, VTCI 没有标准的等级划分, 考虑其取值在一定取值范围内 $[0, 1]$, 可将其取值范围平分为若干区间来确定其状态空间。对于区间数目的确定, 如果区间数目太少, 则预测出的结果精度较差; 但如果区间数目太多, 计算需要的时间比较长。本研究中将取值范围内 $[0, 1]$ 分为 20 个区间来确定马尔可夫链的状态空间。

(2) 根据状态空间确定 VTCI 时间序列中各值所处的状态, 然后计算不同滞时的马尔可夫模型的转移概率矩阵, 该矩阵决定了 VTCI 值状态转移过程的概率法则。

(3) 对 VTCI 序列进行马尔可夫性检验, 若通过检验, 则说明所选 VTCI 时间序列是马尔可夫链, 可应用马尔可夫模型进行分析预测。

对离散序列的马氏性检验通常可用 χ^2 统计量来进行。用 f_{ij} 表示序列中从状态 i 经过一步转移到状态 j 的频数, $i, j \in E$ 。将转移频数矩阵的第 j 列之和除以各行各列的总和所得的值称为边际概率, 记为 $P_{\cdot j}$ 。

$$P_{\cdot j} = \sum_{i=1}^N f_{ij} / \sum_{i=1}^N \sum_{j=1}^N f_{ij}, \text{ 且 } P_{ij} = f_{ij} / \sum_{j=1}^N f_{ij} \quad (2)$$

式中, N 为状态空间数。当序列长度 n 充分大时, 计算统计量

$$\chi^2 = 2 \sum_{i=1}^N \sum_{j=1}^N f_{ij} \left| \log \frac{P_{ij}}{P_{\cdot j}} \right| \quad (3)$$

该统计量服从自由度为 $(N-1)^2$ 的 χ^2 分布, 计算后得统计量 χ^2 值。给定显著性水平 α 为 0.05, 查表可得分位点 $\chi^2_{\alpha}((N-1)^2)$ 的值。若 $\chi^2 > \chi^2_{\alpha}((N-1)^2)$, 则可认为序列 X_t 符合马尔可夫性, VTCI 序列通过马氏性检验。

(4) 计算各阶自相关系数 r_k 。

$$r_k = \sum_{i=1}^{n-k} (X_i - \bar{X})(X_{i+k} - \bar{X}) / \sum_{i=1}^n (X_i - \bar{X})^2 \quad (4)$$

式中, k 为滞时(步长), $k = 1, 2, \dots, m$, m 为按预测需要计算到的滞时最大阶数。 r_k 为第 k 阶滞时的自相关系数, X_i 为 VTCI 序列中第 i 时段的值, \bar{X} 为序列均值, n 为序列的长度。

计算各种滞时的马尔可夫模型的权重 w_k 。

$$w_k = |r_k| / \sum_{k=1}^m |r_k| \quad (5)$$

(5) 分别以预测时段前面若干滞时的 VTCI 所处状态为初始值, 结合其相应的转移概率矩阵, 得到所预测时段 VTCI 的状态概率 $p_i^{(k)}$ ($i \in E$)。

(6) 将同一状态各预测概率加权和作为 VTCI 值处于该状态的预测概率 p_i 。

$$p_i = \sum_{k=1}^m w_k p_i^{(k)} \quad i \in E \quad (6)$$

(7) 预测时段的状态特征值 S 的确定。其思想是将最大概率和其它概率的影响进行综合考虑, 更好地确定预测时段所处的状态。

$$S = \sum_{i=1}^N i \times (p_i^{\eta} / \sum_{i=1}^N p_i^{\eta}) \quad i \in E \quad (7)$$

式中, η 为调整因子, 其值越大, 最大概率的作用越大。 η 取值的方法为: 从步骤 5) 计算出的各状态预测概率值中找出最大 (p_{\max}) 和次大 ($p_{\text{second max}}$) 的概率值, 计算二者的比值 $A = p_{\max} / p_{\text{second max}}$, 最后确定 η 的值。

$$\eta = \begin{cases} 4 & A > 4 \\ A & 1 \leq A \leq 4 \\ 1 & A < 1 \end{cases} \quad (8)$$

(8) 依据 $|S - i| < 0.5, i \in E$ 的准则确定该时段指数所处的预测状态 i 。然后取预测状态取值范围的中间值作为预测的 VTCI 值。

2 结果和分析

本研究中将关中平原 1999 年 5 月上旬至 2003 年 5 月中旬的 VTCI 监测数据作为建模数据, 2003 年 5 月下旬和 6 月上旬的 VTCI 监测数据作为检验数据。根据 1.2 节的方法, 首先对 VTCI 监测数据中每个空间像点的时间序列进行马尔可夫性检验。给定

显著性水平 α 为 0.05, 查表得分位点 $\chi^2_{\alpha}((N-1)^2)$ 的值为 66.054。若 $\chi^2 > 66.054$, 则可认为序列 X_t 符合马尔可夫性, VTCI 序列通过马氏性检验。检验结果见图 1。图中显示除个别像点之外(区域中白色部分), 大部分像点(区域中黑色部分)的时间序列都通过了马尔可夫性检验。考虑不通过马尔可夫性检验的像点极少, 忽略这些像点的非马尔可夫性, 对整个空间像点进行马尔可夫模型 2 步预测。

依照 1.2 节的步骤得到 2003 年 5 月下旬和 6 月上旬的 VTCI 预测结果。图 2 和图 3 分别是 2003 年 5 月下旬、6 月上旬的监测影像和预测影像及其相应的频数分布图。

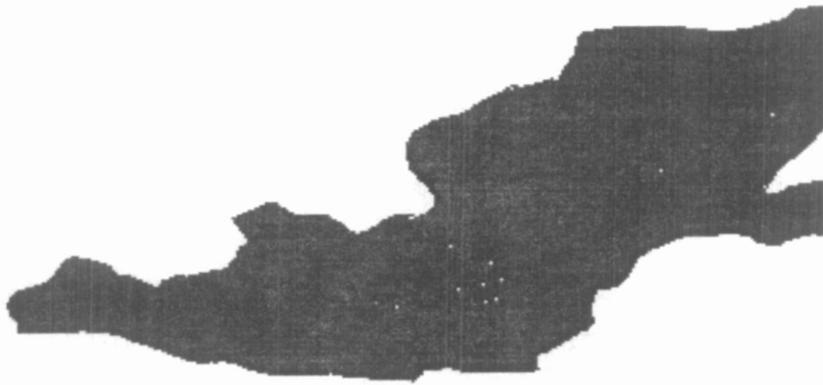


图 1 马尔可夫性检验
Fig.1 Testing for fitting to Markov model

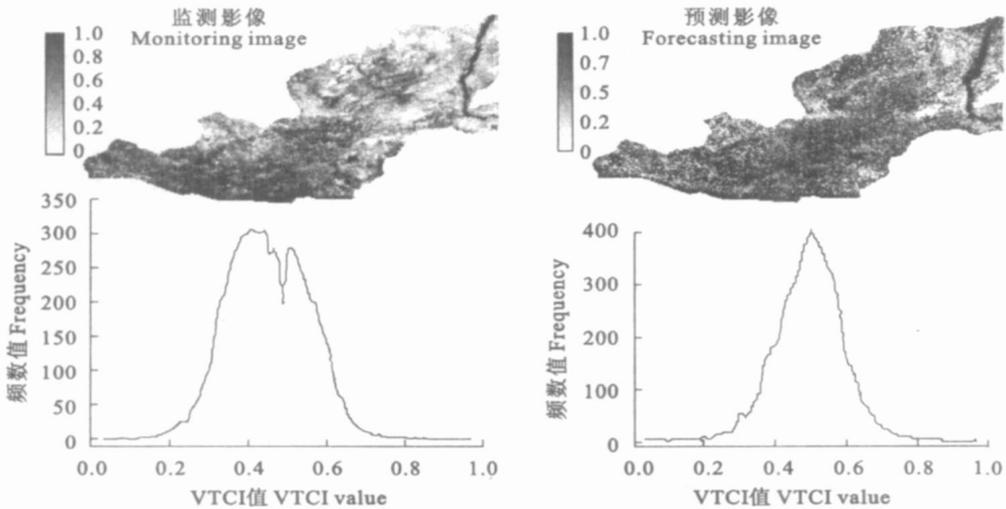


图 2 2003 年 5 月下旬 VTCI 监测影像、预测影像及其频数分布

Fig.2 Monitoring, forecasting images and their own frequency distributing for the last ten-day of May in 2003

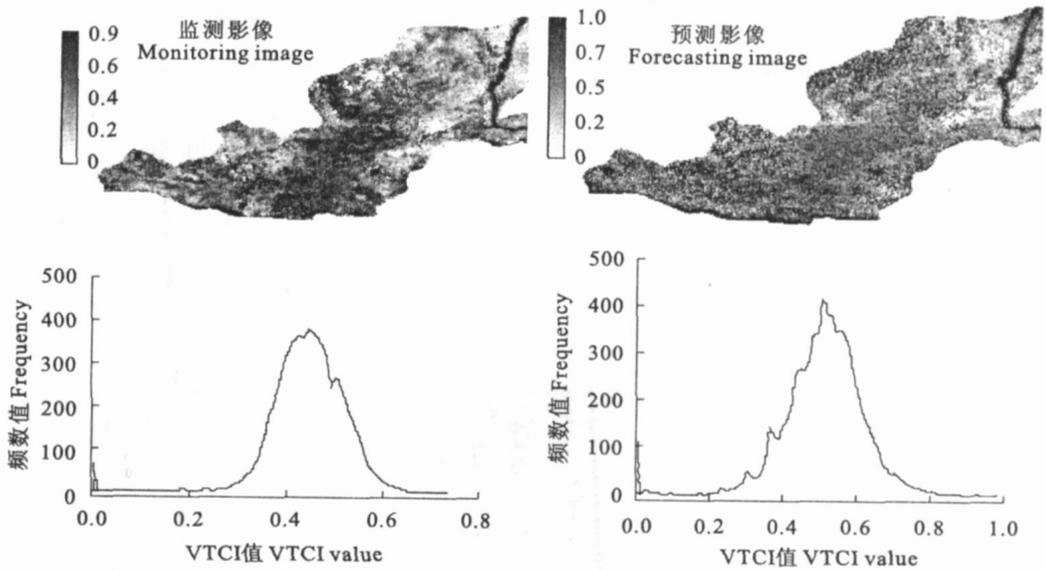


图 3 2003 年 6 月上旬 VTCl 监测影像、预测影像及其频数分布

Fig. 3 Monitoring, forecasting images and their own frequency distributing for the first ten-day of June in 2003

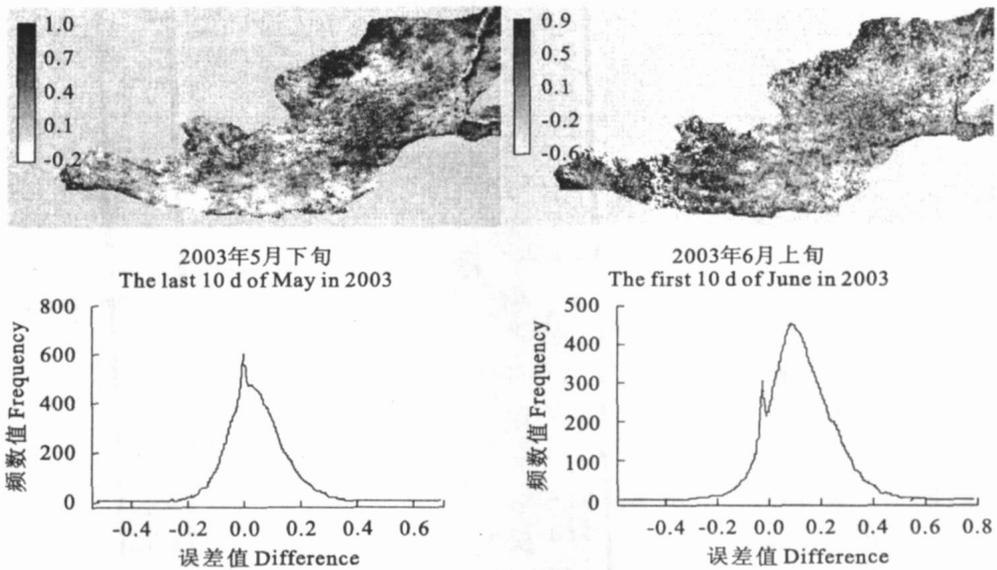


图 4 预测误差及其频数分布

Fig. 4 Difference images and their own frequency distributing

将 2003 年 5 月下旬、6 月上旬实际监测的 VT-Cl 影像和预测影像进行比较,从总体分布上来看,2 个预测影像都基本反映了监测影像的特征。关中平原北部的干旱程度较南部严重,黄河两岸存在着相对较重的干旱,位于晋陕豫三省交界处的黄河(影像右上角)在预测图中界线清晰。但 6 月上旬预测影像在反映监测影像的细节上较 5 月上旬的稍差。相应的频数分布对比可以看出,1 步预测(2003 年 5 月下旬)影像与其相应的监测影像有着相似分布。频数大于 50 的 VTCl 范围相似,监测影像是 [0.27,

0.65], 预测影像是 [0.30, 0.65]。2 步预测(2003 年 6 月上旬)影像与其相应的监测影像的相似性稍差一些,频数大于 50 的 VTCl 范围,监测影像是 [0.33, 0.57], 预测影像是 [0.30, 0.67]。较大不同的是最大频数对应的 VTCl 值,2003 年 5 月下旬相差 0.1 左右,2003 年 6 月上旬相差 0.05。

图 4 是误差影像(预测影像减监测影像)及其频数分布。图 4 的误差分析可以看出,2003 年 5 月下旬的预测误差主要(频数大于 50)在 [-0.15, 0.25] 之间,最大频数误差为 0;2003 年 6 月上旬的预测

误差主要(频数大于 50)在 $[-0.1, 0.4]$ 之间,最大频数误差为 0.1 左右。

表 1 是预测影像和监测影像及其相应误差影像进一步的统计分析。表中可以看出 2003 年 5 月下旬和 6 月上旬的预测影像和监测影像中 VTCI 的最小值、最大值、平均值和标准偏差十分接近。误差影像的统计分析中 2003 年 5 月下旬预测误差平均值

为 0.042, 6 月上旬误差平均值为 0.152, 显示 2003 年 5 月下旬预测精度优于 6 月上旬。

由以上的分析可以看出,2003 年 5 月下旬和 6 月上旬的预测基本反映了实际监测的干旱状况,预测精度较好。其中 2003 年 5 月下旬的 1 步预测好于 6 月上旬 2 步预测的结果。

表 1 监测、预测及其误差影像的统计分析

Table 1 Statistics for monitoring, forecasting and difference images

VTCI 影像 VTCI images	最小值 Minimal value	最大值 Maximal value	平均值 Mean value	标准偏差 Standard deviation
2003 年 5 月下旬 The last ten-day of May in 2003	监测 Monitoring	0	0.984	0.449
	预测 Forecasting	0	0.979	0.489
	误差 Difference	-0.539	0.715	0.042
2003 年 6 月上旬 The first ten-day of June in 2003	监测 Monitoring	0	0.812	0.448
	预测 Forecasting	0	0.978	0.499
	误差 Difference	-0.595	0.864	0.152

3 结 论

本文利用加权马尔可夫模型对陕西关中平原的 VTCI 时空序列进行预测,将 VTCI 的取值范围分为 20 个区间,以此确定马尔可夫状态空间,最后通过加权马尔可夫模型方法预测出每个像点的 VTCI 值。预测结果显示 2 步预测的预测影像和监测影像都较为相似,频数分布图也显示较好的相似性,进一步的误差分析表明预测误差较小,2 步预测的平均误差分别为 0.042 和 0.152。因此加权马尔可夫模型方法可以用于关中平原 VTCI 数据的 2 步预测,以评估关中平原的干旱状况。该方法也为具有时空变化特征的数据序列预测提供了可借鉴的方法。

参 考 文 献:

- [1] 宋连春,邓振镛,董安祥,等.干旱[M].北京:气象出版社,2003.
- [2] Boken V K, Cracknell A P, Heathcote R L. Monitoring and predicting agricultural drought—a global study[M]. New York: Oxford University Press, 2005. 57—87.
- [3] 王鹏新,龚健雅,李小文.条件植被温度指数及其在干旱监测中的应用[J].武汉大学学报(信息科学版),2001,26(5):412—418.

- [4] 王鹏新, Wan Zhengming, 龚健雅, 等.基于植被指数和土地表面温度的干旱监测模型[J].地球科学进展,2003,18(4):527—533.
- [5] 孙威,王鹏新,韩丽娟,等.条件植被温度指数干旱监测方法的完善[J].农业工程学报,2006,22(2):22—26.
- [6] Wan Z, Wang P, Li X. Using MODIS land surface temperature and normalized difference vegetation index products for monitoring drought in the southern Great Plains, USA [J]. International Journal of Remote Sensing, 2004, 25(1):61—72.
- [7] Paulo A A, Ferreira E, Coelho C, et al. Drought class transition analysis through Markov and Loglinear models, an approach to early warning[J]. Agricultural Water Management, 2005, 77(1—3):59—81.
- [8] Wilks D K. Interannual variability and extreme value characteristics of several stochastic daily precipitation models[J]. Agricultural and Forest Meteorology, 1999, 93:153—169.
- [9] 孙才志,张戈,林学钰.加权马尔可夫模型在降水丰枯状况预测中的应用[J].系统工程理论与实践,2003,18(4):101—106.
- [10] 温克刚,翟佑安.中国气象灾害大典(陕西卷)[M].北京:气象出版社,2005.
- [11] 陕西历史自然灾害简要纪实编委会.陕西历史自然灾害简要纪实[M].北京:气象出版社,2002.
- [12] 汪荣鑫.随机过程[M].西安:西安交通大学出版社,1988.
- [13] 易丹辉.统计预测——方法与应用[M].北京:中国统计出版社,2001.

- [J]. 地理学报, 2005, 60(1), 3-11.
- [12] 徐影, 丁一汇, 李栋梁. 青藏地区未来百年气候变化[J]. 高原气象, 2003, 22(5), 451-457.
- [13] 王绍武, 龚道溢. 对气候变化问题争议的分析[J]. 地理研究, 2001, 20(2), 153-160.
- [14] 吕晓蓉, 吕晓英. 青藏高原东北部草地气候暖干化趋势分析[J]. 中国草地学报, 2002, 24(4), 9-14.
- [15] 雷晓云, 陈惠源, 张瑞民. 干旱地区地表水资源预测的多层递阶组合模型[J]. 灌溉排水, 1998, 17(4), 26-29.
- [16] 黄嘉佑. 气象统计分析与预报方法[M]. 北京: 气象出版社, 2002. 81.
- [17] 黄忠恕. 波谱分析方法及其在水文气象学中的应用[M]. 北京: 气象出版社, 1983. 6-93.
- [18] 李林, 王振宇, 秦宁生, 等. 环青海湖地区气候变化及其对荒漠化的影响[J]. 高原气象, 2002, 21(1), 59-65.
- [19] 颜亮东, 张国胜, 李凤霞. 我国北方牧区天然牧草产量形成的气候模式[J]. 草业科学, 2003, (7), 8-11.
- [20] 颜亮东, 李凤霞, 张国胜. 青海省环湖地区天然牧草返青与气象条件的关系及其预报[J]. 中国农业气象, 1998, (2), 42-45.
- [21] 郭连云, 张旭萍, 公保才让, 等. 共和盆地天然草地牧草产量与水、热因子的灰色关联度分析[J]. 中国草食动物, 2007, 27(3), 6-9.
- [22] 范青慈, 李希来. 青海高寒草甸草地生产力与气候因素的灰色关联度分析[J]. 草业科学, 2003, 20(3), 8-11.

Characteristics of the change of climate factors affecting grass growth in the natural grassland of Gonghe basin

GUO Lianyun¹, ZHANG Xuiping¹, DING Shengxiang²

(1. Hainan Prefectural Meteorological Bureau, Gonghe, Qinghai 813000, China;

2. Tongde County Meteorological Bureau, Tongde, Qinghai 813210, China)

Abstract: Analysis is made of the tendency, stage and periodicity of the change of climate factors affecting grass growth in the natural grassland of Gonghe basin in the past 47 years. The results showed that, among the various factors, the average temperature and accumulated temperature from April to September presented a significant increasing tendency and cyclical stages; while the accumulated temperature, precipitation, sunshine hours and precipitation in the previous autumn showed a significant oscillating cycle, with the total sunshine hours having 2~6 a oscillating cycle and precipitation in the previous autumn having 1~2 a and 5~6 a oscillating cycles.

Keywords: natural grassland; climate factor; climate change; Gonghe basin

(上接第 200 页)

Drought forecasting based on the vegetation temperature condition index using weighted Markov model

HAN Ping¹, WANG Pengxin^{2*}, WANG Jiahui¹, SUN Wei², ZHU Dehai²

(1. College of Science, China Agricultural University, Beijing 100094, China;

2. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

Abstract: The spatial and time series of vegetation temperature condition index (VTCI) for Guanzhong Plain and Weibei Tablelands are analyzed using weighted Markov model. The VTCI images of 2 lead-time prediction are got. Comparing the monitoring images with the forecasting images, we can find the similar characters showing similar drought distribution between the monitoring image and the forecasting image. The further difference detection in difference images shows that the average difference of 1-step prediction is 0.042, and the average difference of 2-step prediction is 0.152.

Keywords: vegetation temperature condition index; weighted Markov model; drought forecasting