

运用混沌理论预测粮食产量

姜会飞¹ 温德永² 廖树华¹ 李树岩¹ 吴文良¹

(1. 中国农业大学 资源与环境学院, 北京 100094; 2. 加拿大滑铁卢大学 大气科学中心, 滑铁卢, N2L3G1)

摘要 运用混沌理论的原理和方法预测粮食产量, 探寻一种准确率高、时效长、成本低、便于普及推广的区域粮食产量预测方法。把粮食单产序列分解为趋势产量和气象产量 2 部分, 用 5 年滑动平均法求得趋势产量后, 通过分析粮食单产中气象产量分量的时间序列的变化规律, 论述了由气候变化引起气象产量的年际波动具有混沌特性, 具体表现为平稳的马尔柯夫型状态转移概率和相同的概率密度分布特点。并以陕西省武功县、河北省禹县和涿州 3 个地区 1949—1999 年小麦单产序列为例, 运用 MATLAB7.0 语言编程构建了粮食产量的混沌预测模型, 对 2000 年小麦单产进行预测, 准确率分别为 98.1%、92.1% 和 97.1%。预测结果表明, 粮食产量的混沌预测法相对准确而有效, 准确率在 92% 以上, 可作为粮食产量预测的新方法加以推广。

关键词 气象产量; 时间序列; 混沌理论; 粮食预测; MATLAB

中图分类号 S 165.27

文章编号 1007-4333(2006)01-0047-06

文献标识码 A

Prediction of grain yield using chaos theory

Jiang Huifei¹, Deyong Wen², Liao Shuhua¹, Li Shuyan¹, Wu Wenliang¹

(1. College of Resource and Environment, China Agricultural University, Beijing 100094, China;

2. Waterloo Center for Atmospheric Sciences, University of Waterloo, Waterloo N2L3G1, Canada)

Abstract Analyzing the interannual fluctuation of weather on grain yield, this paper establishes that the weather fluctuation has the Chaos characteristic resulted from climate variability. Based on the Chaos theory and wheat yield of Wugong in Shanxi province, Yuxian and Zhuozhou in Hebei province from 1949 to 1999, the yield predication models of these three districts have been set up. Using these models the predication veracity of wheat yield in the three districts in 2000 are 98.1%, 92.1% and 97.1%, respectively. The yield prediction veracities are all above 92%. Chaos predication Model of grain yield can be recommended as a new method of yield prediction.

Key words relative weather yield; time series; Chaos theory; grain prediction; MATLAB

预测粮食作物产量的变化趋势对国家粮食安全、政策宏观调控和区域结构调整有着十分重要的意义。目前国际上预测作物产量较为流行的有灰色马尔柯夫模型预测法、遥感技术预测法、统计动力学生长模拟法和系统综合因素预测法 4 种^[1-3]。

灰色马尔柯夫模型结合灰色模型预测趋势产量和马尔柯夫模型预测气象产量的方法, 比较合理和准确^[2-3]。该模型把气象产量随时间的变化当作随机过程来处理, 并以最大转移概率预测状态, 预测总体水平提高了, 但却忽略了小概率事件发生的可能

性, 在小概率事件发生的年份, 预测误差较大。

遥感估产具有快速、宏观、动态等优点, 是大范围田面平整、种类单一的作物长势监测和产量预测的有效手段^[4]。但是, 遥感估产受到分辨率、时相等因素的影响, 对地形复杂、作物种类多、面积狭小的农田产量预测成本较高。

统计动力学生长模拟在区域参数化的基础上动态模拟作物生长进程, 反映产量形成机理, 其预测效果受气候预测水平制约, 难于及时获得大面积的各种数据, 该方法仍处于小范围实验阶段^[1], 难以满

收稿日期: 2005-11-10

基金项目: 国家“十五”科技攻关资助项目 (2001BA513B3-1)

作者简介: 姜会飞, 副教授, 在读博士, E-mail: jianghuifei@sohu.com; 吴文良, 教授, 博士生导师, 通讯作者, 主要从事农业生态与政策研究, E-mail: wuwenl@cau.edu.cn

足大面积较长时效的产量预测要求。

系统综合因素预测充分考虑气象条件、社会经济和科技水平等各种因素的影响,在作物产量预测方面取得了精度高、预测时效长等成果^[1]。但这种方法要求的基础数据量大、计算复杂,难以在数据缺乏的区域普及。

本研究旨在用混沌理论探寻预测准确率较高、时效相对长、成本低、便于普及推广的区域粮食产量预测的新方法。

1 混沌理论及其应用简介

美国著名的气象学家洛仑兹在数值实验中首先发现混沌现象^[5-6],认为混沌现象的随机和无规则行为来源于确定性和有序性,它具有貌似随机却并非随机的特性。随着对混沌现象的深入研究,混沌理论革新了科技预测理论和方法体系,运用该理论解决了众多一般理论难以解决的数学、物理、化学和生物等学科中的问题,并取得了科技界瞩目的成果^[7]。

气象学家们将混沌理论应用于气候系统中,发展成为混沌气候学。在过去20多年,混沌理论应用到地球物理学的各个领域^[8-11],大大地提高了天气气候预测和海洋潮汐预报的准确率。

气候变化具有混沌特性,在表面看来是无序的、随机的,但是在气体初始状态条件下按气体状态方程变化规律发展的,是有规律的、有序的^[5]。作物气象产量的年际变化是气候变化的混沌过程所致,也应该具有混沌特性。卢志光通过分析北京市粮食产量和降水量的相关关系,以粮食减产率20%的年份界定旱灾年,运用混沌理论成功地预测了北京旱灾的发生情况,准确率达到89%^[6],开创了混沌理论应用于农业气象预测的先河。

本研究拟探寻气象产量波动内在规律的混沌机制,建立粮食产量的混沌预测模型,进行将混沌理论应用到农业气象产量预测的尝试。

2 资料来源及处理

小麦实际单产序列为52年(1949—2000),3个序列资料分别从陕西武功县、河北禹县和涿州统计部门获得。

作物单产(y_a)常被分解为时间趋势产量(y_t)和气象波动产量(y_w)^[1-3,12-13]。以5年滑动平均法求算逐年(1951—1997)小麦单产中的趋势产量,实际

单产与趋势产量之差即为气象波动产量。气象波动产量与趋势产量之比称为相对气象产量。相对气象产量为正值表示当地气象条件对小麦生产有利,增产;负值表示气象条件对生产不利,减产。以 y_w 表示相对气象产量百分率(%),即:

$$y_w = \frac{y_a - y_t}{y_t} \times 100\% \quad (1)$$

从3个地区小麦单产、趋势产量和相对气象产量百分率(%)的年变化曲线图(图1)可以看出:趋势产量变化相对稳定,总体上呈现随时间逐渐上升的趋势。这是因为农业科学技术和管理水平提高使农作物的单产水平随之提高。而气象产量出现较大的年际波动,常被当作随机过程来处理^[2-3,12-13]。这是由于引起气象产量年际波动的气象条件具有看似随机却并非随机的混沌特性所致。因此,气象产量的时间序列也不是无序的随机过程,应该存在着某种内在的规律性。目前,国内外的大量研究普遍认为气候因子和气象产量的时间序列具有马尔柯夫过程的特点^[2-3,12-14]。

3 马尔柯夫随机过程

若随机过程 $x(t)$,在时刻 t 系统的状态以 E 表示,对于时刻($> t$)系统所处状态与时间 t 以前所处状态无关,这一过程称为马尔柯夫过程。马尔柯夫过程又称为无后效的随机过程。在这个随机过程中,每一时刻都对应着一种状态,而系统可能存在的状态数目却是有限的,因此,马尔柯夫随机过程通常看作是由若干个可以相互转换的状态构成的时间序列,也可称为马尔柯夫链。如果在这个时间序列中,状态之间的转移概率不随时间而变,这样的过程称为平稳马尔柯夫过程。在天气气候模拟和农业气象产量预报的应际应用中,气候要素时间序列及其受气象条件的年际变化引起的产量波动随机过程常当作平稳马尔柯夫过程来处理^[2,14]。

在1个马尔柯夫链中,系统状态的转移可以是系统中任一状态转移至另一状态或与自身相同的状态,假设系统共有 m 种状态,记为 $E_i (i=1, 2, \dots, m)$ 。对于出现其中任一状态的事件构成一个完备事件群,随时间由从状态 E_i 转移到状态 E_j 事件概率称为转移概率,记为 $p_{ij} = P(E_j/E_i)$,表示由前一状态 E_i 条件下事件 E_j 发生的概率(随机可能性)。

由各状态的各种转移概率为元素组成的矩阵称为转移概率矩阵,记为:

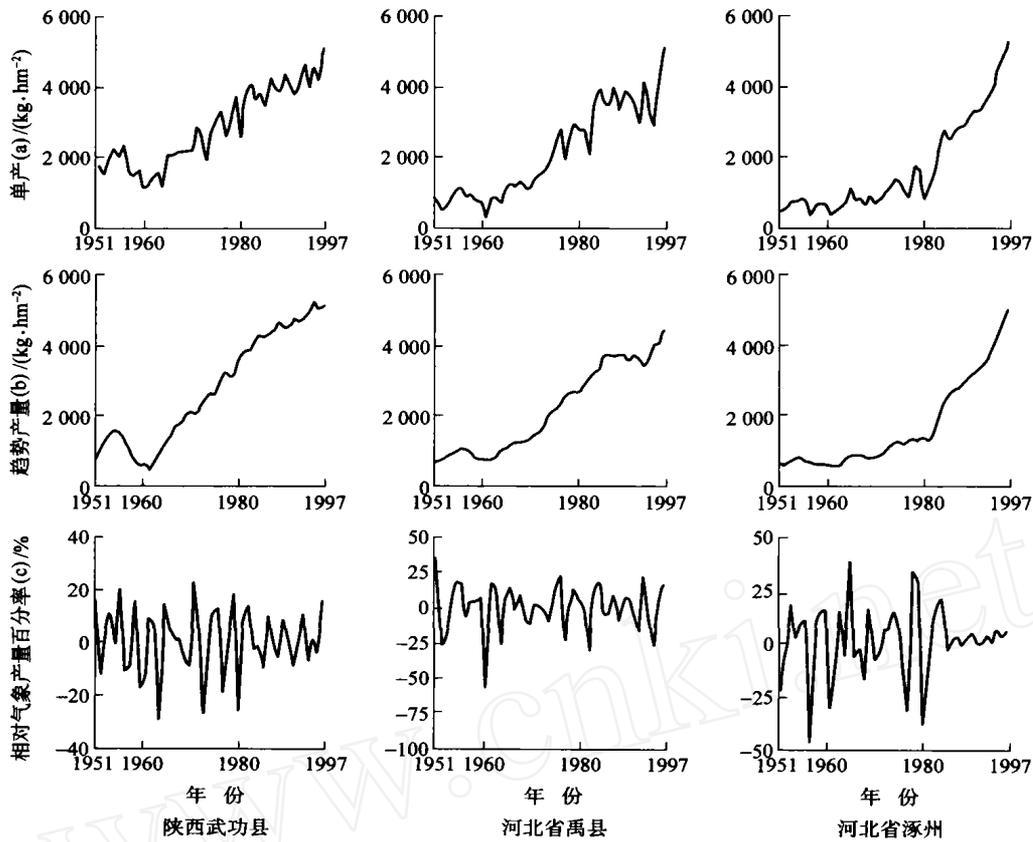


图 1 3 地区小麦单产、趋势产量和相对气象产量百分率的年变化曲线

Fig. 1 Yearly curve of wheat yield (a), trend yield (b) and relative weather yield (c) in three regions

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1m} \\ p_{21} & p_{22} & \dots & p_{2m} \\ \dots & \dots & \dots & \dots \\ p_{m1} & p_{m2} & \dots & p_{mm} \end{bmatrix} \quad (2)$$

其中, $0 \leq p_{ij} \leq 1, \sum_{j=1}^m p_{ij} = 1$ 。如果时间序列中样本数量有限,可以近似地用状态转移频率 \hat{p}_{ij} 来代替转移概率 p_{ij} ,即可用转移频率矩阵

$$\hat{P} = \begin{bmatrix} \hat{p}_{11} & \hat{p}_{12} & \dots & \hat{p}_{1m} \\ \hat{p}_{21} & \hat{p}_{22} & \dots & \hat{p}_{2m} \\ \dots & \dots & \dots & \dots \\ \hat{p}_{m1} & \hat{p}_{m2} & \dots & \hat{p}_{mm} \end{bmatrix} \quad (3)$$

并把它作为概率矩阵的估计^[15]。

4 随机过程的分布与轨迹

在马尔柯夫链中,状态对应的可能是一个数值区间而不是数据点,即与状态对应的数值不是唯一的。假设随机变量 x 在随时间变化而发生的随机过程有图 2 所示的 4 种时间序列用来描述随机过程

的 3 种状态 E_i 、 E_j 和 E_k 分别对应着数值区间为 $(x_{i1}, x_{i2}]$ 、 $(x_{j1}, x_{j2}]$ 和 (x_{k1}, x_{k2}) ; 随机过程 1 中,随机变量 x 在取值空间 (x_0, x_n) 内所有数值点集上出现的概率都等于 p_0 ,这个随机过程的概率密度曲线则可以用直线 AB 来描述;随机过程 3 中,变量 x 在 (x_0, x_n) 数值点集上出现的概率不相等,但在同一状态 (E_i 、 E_j 或 E_k) 区间内出现的概率是相等的,在状态 E_i (对应线段 CD) 内的变量出现的概率等于 p_2 ,在状态 E_j (对应线段 EF) 内的变量出现的概率等于 p_5 ,在状态 E_k (对应直线 GH) 的变量出现的概率等于 p_4 。利用转移概率预测到下一步变量最可能出现的状态,而变量在同一状态区间内出现的概率是相同的,也就是说变量发生在这一状态区间内的任意数值都具有同等的机会。也就是说,变量出现在状态上限 (对应最大值)、状态下限 (对应最小值) 和区间任意值的概率是相等的。对于随机过程 1 和 3 转移状态的平均值可取中值。但随机过程 2 和 4 中同一状态区间中不同数值点出现的概率是变化的,在下一步转移状态中概率大的数值点出现的可能性较大,变量平均值就倾向状态区间中概率较

大的数值点。因此,对于随机过程状态平均值的求算不能用简单取上下限平均的方法,我们认为用下面的方法是较合理的。

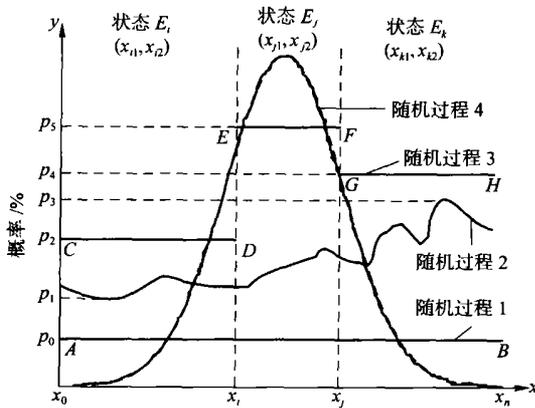


图2 几种随机过程概率密度函数曲线

Fig. 2 Four probability consistency curves of stochastic series

假设已知随机变量概率密度分布函数为 $f(x)$, 运用模型随机生成具有此分布型的序列样本数 N (本文取 1 000) [16]。如果在随机生成的由 N 样本组成的序列中, 产生在状态 E_i , 状态变量 x 取值范围为 $x_i [x_{i1}, x_{i2}]$ 的变量数为 M , 分别记为 $x_1, x_2, x_3, \dots, x_m$ 。因此, 在状态区间变量平均值为

$$\bar{x}_i = \frac{\sum_{j=1}^m x_j}{M} \quad (4)$$

利用马尔柯夫随机模型从某时刻的初始状态可以预测到下一时刻最大可能出现的状态, 这个状态对应的可能是一个数值区间, 并不是具体的数据点值。在以往的许多研究中, 常取区间中间点作为预测平均值 [23]。笔者认为, 转移概率预测的是最大可能出现的状态, 那么预测值可能是状态区间上的任意数值点。随机变量 x 在 m 种状态的任一状态区间出现的概率机会均等, 则可以取区间中间点作为平均值, 但如果在状态区间各数值点出现的概率机会不均等, 则平均值应取状态区间各数值点的加权平均。因此, 运用转移概率预测状态区间后, 确定预测值时要事先弄清楚随机过程的概率分布类型, 以及构成随机过程的各状态在其对应的数值区间上的概率密度情况。

研究表明: 影响气象产量的各种气象因子的时间序列具有正态分布的特征 [18], 即极端气象条件 (通常表现为致灾因子) 发生的概率较小, 一般气象

条件发生的概率较多。在粮食作物生产中, 相对气象产量的时间序列也普遍表现出正态分布特征 [17-18], 可以认为相对气象产量百分率 y_w (%) 就是这样一随机变量 x , 其对应的概率密度 $f(x)$ 为

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2} \quad (5)$$

式中: μ 为时间序列平均值, σ^2 为序列的均方差。在状态 E_i , 变量 x 对应取值范围为 $x_i [x_{i1}, x_{i2}]$, 变量 x 出现在该状态的概率为 [17]

$$\int_{x_{i1}}^{x_{i2}} f(x) dx = p(x_{i1} < x < x_{i2}) \quad (6)$$

5 产量预测模型及检验

相对气象产量序列是受气候条件年际变化引起的产量波动, 常被看作是随机事件。这种随机过程 $y_w(t)$ 是气候变化的混沌过程所致。其规律性表现为具有平稳的状态转移概率和相同的概率密度分布特点, 也就是说这种随机过程具有特定的轨迹。这种随机过程称为随机轨迹, 其轨迹由平稳的状态转移概率和相同的概率分布函数共同产生。因此, 在已知初始状态和随机轨迹的情况下, 构建气象产量混沌发生器, 预测气候波动产量。

假设随机序列共有 m 种状态 $E_i (i = 1, 2, \dots, m)$, 变量 $x(t)$ 初始状态为 E_j , 变量由从状态 E_j 转移到状态 E_i 事件概率为 p_{ji} 。联立式 (3) ~ (6), 得到预测变量 $x(t)$ 的在下一时刻的混沌解:

$$x(t+1) = \sum_{i=1}^m [x_i(t+1) \times p_{ji}] \quad (7)$$

以小麦相对气象产量百分率 (%) y_w 为指标, 把 y_w 的时间序列分为 3 种年型: $y_w < -5.0\%$ 为灾年, $-5.0\% < y_w < 5.0\%$ 为平年, $y_w > 5.0\%$ 为丰年 [3]。相应地, 3 种年型对应的 3 种状态分别以 E_1 、 E_2 和 E_3 表示。本研究以陕西武功、河北禹县和涿州 3 个地区 1949—2000 年小麦产量序列为例, 把单产分解为 1951—1997 年的趋势产量和相对气象产量序列, 利用马尔柯夫模型得到 3 个地区的状态转移概率分布 (图 3)。

根据 3 个地区的相对气象产量百分率 (%) 的 μ 和 σ 值, 确定其概率密度分布函数分别为

陕西省武功县:

$$y = \frac{1}{12.5 \sqrt{2}} e^{-\frac{1}{2} \left(\frac{y-0.5}{12.5} \right)^2} \quad (8)$$

河北省禹县：

$$y = \frac{1}{16.7 \sqrt{2}} e^{-\frac{1}{2} \left(\frac{x+0.1}{16.7} \right)^2} \quad (9)$$

河北省涿州：

$$y = \frac{1}{17.0 \sqrt{2}} e^{-\frac{1}{2} \left(\frac{x-1.6}{17.0} \right)^2} \quad (10)$$

绘制 3 个地区相对气象产量百分率 (%) 概率密度函数曲线见图 4。

分别建立以上 3 个地区 1951—1997 年的相对气象产量的马尔柯夫模型和概率密度分布模型的联合方程组, 求解得 1998 年的相对气象产量 (%) 预测值。根据趋势产量为连续 5 年实际单产 y_a 的滑动

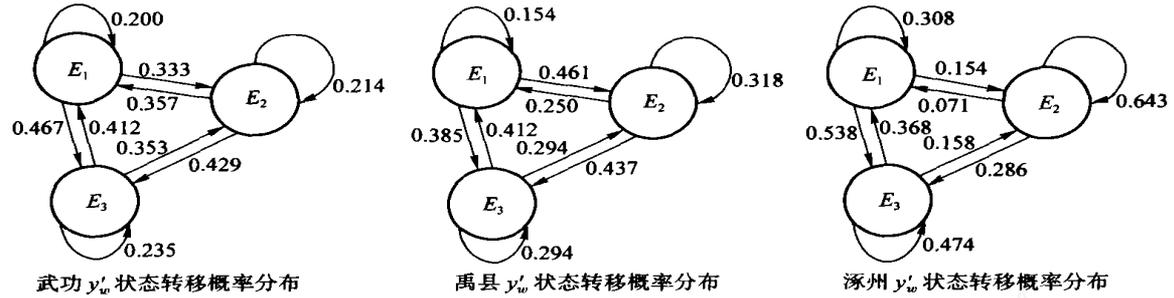


图 3 3 地区转移概率分布图

Fig. 3 Three simple Markov chain transfer probability diagram

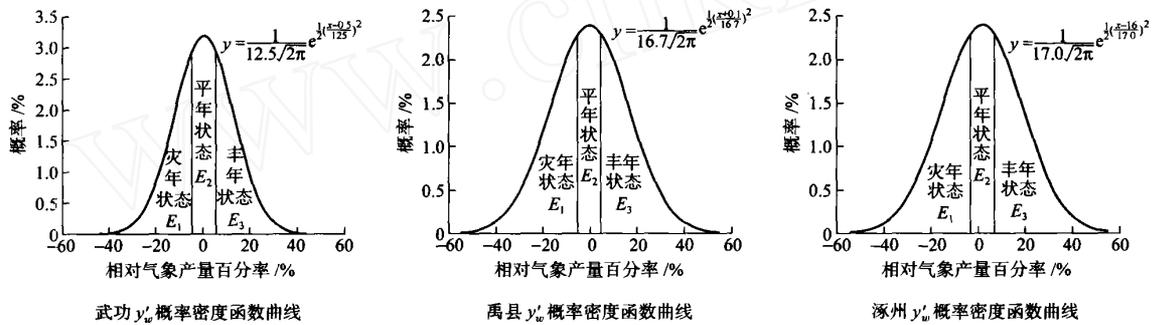


图 4 3 地区相对气象产量概率密度函数曲线

Fig. 4 Three probability consistency curve of relative weather yield

平均值, 趋势产量的预测年 (1998 年) 在 5 年序列中排列第 3, 产量序列 $y_{a1}, y_{a2}, y_{a3}, y_{a4}, y_{a5}$ 对应着 1996—2000 年的小麦单产, 而其中 1996—1999 年单产是既定的, 2000 年的小麦产量 y_{a5} 是要预测的。用式 (1) 描述 1998 年的相对气象产量如下：

$$y_w = \frac{y_{a3} - y_t}{y_t} \times 100\% = \frac{y_{a3} - \sum_{i=1}^5 y_{ai}/5}{\sum_{i=1}^5 y_{ai}/5} \times 100\% = \frac{5y_{a3} - \sum_{i=1}^4 y_{ai} - y_{a5}}{\sum_{i=1}^5 y_{ai} + y_{a5}} \times 100\% \quad (11)$$

上式 (11) 经转换得到 2000 年产量

$$y_{a5} = \frac{500 y_{a3} - (100 + y_w) \sum_{i=1}^4 y_{ai}}{100 + y_w} \quad (12)$$

表 1 小麦产量混沌预测结果与实况对照表

Table 1 Result of prediction and actual wheat yield

地区	预测产量	实际产量	准确率/ %
	(kg · hm ⁻²)		
河北禹县	4 687.9	4 345.0	92.1
河北涿州	5 409.4	5 257.5	97.1
陕西武功	4 250.5	4 171.0	98.1

预测产量与实际产量较为接近, 产量预测准确率均在 92% 以上。因此, 运用混沌理论预测作物产量的方法是有效的, 其预测结果基本可以达到应用

要求。

6 讨论

农作物气候产量预测和农业气候灾害预测是混沌气候学在农业应用领域的分支。对于这些累积效果明显的混沌事件的预测,可以用高阶转移概率辅助预测,以进一步提高预测的准确率。

一般地,大多数来自实际生产的时间序列,特别是描述气候过程及其影响的时间序列数量是很有限的,远不及理论判据所要求的完备事件的资料数量,这个时间序列理论应用上的“数量瓶颈”问题制约着作物产量序列内部普遍规律的探究,是产量预测误差的主要原因,也是今后需要进一步研究和解决的课题。

另外,本研究对产量状态只分为有限的3种。如果进行状态无限微分,从理论上讲,预测准确率应该更高。但限于计算机的运算速率和性能,实现状态的无限微分是不可能的,但适当细分是可行的。

参 考 文 献

- [1] 陈锡康,杨翠红. 农业复杂巨系统的特点与全国粮食产量预测研究[J]. 系统工程理论与实践,2002,6:108-112
- [2] 张淑娟,何勇. 基于趋势-状态预测方法的粮食产量预测[J]. 浙江大学学报(农业与生命科学版),2001,27(6):673-676
- [3] 陈怀亮,张雪芬,毛留喜. 基于产量阶段的河南省夏玉米灰色-马尔柯夫预测模型[J]. 中国农业气象,1996,17(6):45-48
- [4] 马玉平,王石立,张黎. 基于遥感信息的华北冬小麦区域生长模型及模拟研究[J]. 气象学报,2005,63(2):204-215
- [5] Lorenz E N. 混沌的本质[M]. 北京:气象出版社,1997:23,67-68
- [6] 卢志光,白丽萍,卢丽. 运用混沌理论制作长期灾害预报模型初探[J]. 中国农业大学学报,2002,7(3):43-46
- [7] Vary Coates, Mahmud Farooque, Richard Klavans, et al. On the future of technological Forecasting [J]. Technological Forecasting and Social Change,2001,67:1-17
- [8] Sivakumar B. Chaos theory in geophysics: past, present and future[J]. Chaos, Solitons and Fractals,2004,19:441-462
- [9] Sannasiraj S A, Hong Zhang, Vladan Babovic, et al. Enhancing tidal prediction accuracy in a deterministic model using chaos theory[J]. Advances in Water Resources,2004,27:761-772
- [10] Hunt J C R. Environmental forecasting and turbulence modeling[J]. Physica D,1999,133:270-295
- [11] Sivakumar B. Forecasting monthly streamflow dynamics in the western United States: a nonlinear dynamical approach [J]. Environmental Modelling & Software,2003,18:721-728
- [12] Matis J H, Saito T, Grantt W E. A markov Chain approach to crop yield forecasting [J]. Agricultural Systems,1985,18:171-187
- [13] Matis J H, Birkett T, Boudreaux D. An application of the markov Chain approach to forecasting cotton yields from surveys[J]. Agricultural Systems,1989,29:357-370
- [14] Jones P G, Thornton P K. Spatial and temporal variability of rainfall related to a third-order Markov Model [J]. Agricultural and Forest Meteorology,1997,86:127-138
- [15] 黄嘉佑. 气象统计分析预报方法[M]. 北京:气象出版社,2004:239-242
- [16] Cameron D S, Beven K J, Tawn J, et al, Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty) [J]. Journal of Hydrology,1999,219:169-187
- [17] 霍治国,李世奎,王素艳,等. 主要农业气象灾害风险评估技术及其应用研究[J]. 自然资源学报,2003,18(6):692-703
- [18] Daryl Mummery, Michael Battaglia. Significance of rainfall distribution in predicting eucalypt plantation growth, management options, and risk assessment using the process-based model CABALA [J]. Forest Ecology and Management,2004,193:283-296