

三江平原地区粮食产量预报模型研究

刘振忠 徐梅

(黑龙江八一农垦大学)

摘要: 运用聚类分析方法依据农业气候条件对三江平原地区进行分类,在此基础上,选取玉米、大豆、小麦三种作物产量的预报点,建立了作物产量的定性和定量预报模型,即年景趋势预报模型、逐步回归周期分量预报模型和多层递阶预报模型。统计检验和实际预报结果表明所得到的预报模型是有实际意义的。

关键词: 粮食产量; 预报模型; 聚类分析; 年景趋势

1 三江平原地区农业气候分类和预报点的选取

三江平原地区包含 19 个市县,地域间气候差异较大,依据气象因子进行作物产量的预报工作,预报点的选取是要首先解决的问题。首先,我们运用聚类分析方法,依据农业气候条件对该地区以市县为单位进行分类,然后在每一类中分别选取一个县作为该类作物产量的预报点。

考虑气象条件对农业生产的影响,选取的分类指标是:

- 1) 年平均气温;
- 2) > 0 的日数;
- 3) 4~9 月降水量;
- 4) 5~9 月日照时数;
- 5) 10 的积温;
- 6) 干燥指数。

由多元统计分析中的聚类分析方法,设有 $n \times m$ 原始数据阵

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}$$

其中 n 为待分类样品个数, m 为指标个数。首先对原始数据进行标准差标准化变换,以消除数据中由于量纲不同而对分类结果产生的影响,其变换公式为

$$x_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m$$

其中: x_{ij} 为变换后的数据, $s_j = \left[\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \right]^{1/2}$ 称为第 j 个变量的标准差, $\bar{x}_j =$

$\frac{1}{n} \sum_{i=1}^n x_{ij}$ 为第 j 个变量的平均值。样品间的距离采用欧氏距离,聚类采用离差平方和法。

运用计算机求解,并画出聚类谱系图(略),从而将 19 个市县分为三类,其分类结果为:

第一类: 抚远、萝北、饶河、鹤岗。这是三江平原地区气温较低,降水较多,积温较少的一类地区。

第二类:同江、绥滨、桦川、集贤、友谊。这是该地区温度较高,降水较少的干旱地区。

第三类:富锦、佳木斯、双鸭山、汤原、宝清、依兰、桦南、勃利、虎林、密山。这是三江平原地区农业气候条件较好的一类地区。

我们在这三个类别中,分别选择了萝北、友谊、宝清3县作为粮食产量的预报点。预报的作物品种为玉米、大豆、小麦3种作物,定量预报的单位为 $\text{kg}\cdot\text{hm}^{-2}$ 。资料年代为1958年至1994年,共计37年。限于篇幅,本文只介绍萝北县预报站点的工作。

2 作物产量序列的处理

通常认为,作物产量 Y_t 主要由两部分构成,即由农业生产技术水平、土地条件、农业政策等确定的趋势产量 Y_{it} 和当年气象条件所确定的气象产量 Y_{wt} ,即 $Y_t = Y_{it} + Y_{wt} + e_t$,其中 e_t 为误差项,一般可以忽略不计。粮食产量预报工作从对产量资料的处理上,可以分为两个方面,一个是将产量序列本身看作是一个时间序列,利用时间序列分析方法建立产量预报模型;另一个是对产量资料进行处理,模拟出趋势产量,进而得到气象产量,对气象产量进行分析并建立预报模型。预报年的产量即为趋势产量的预报值和气象产量的预报值之和^[1]。

本文以气象产量的预报为主。趋势产量的提取分别使用了滑动平均法、直线滑动平均法和正交多项式法^[1],气象产量由作物的实际产量与趋势产量之差得到。

我们将产量预报分为产量的年景趋势预报(即定性预报)和定量预报两个方面。

2.1 作物产量的年景趋势预报

气象产量是由气象因素确定的那部分产量,是气象条件对作物产量的贡献,因此,依据气象条件对作物产量的趋势预测是符合实际情况的。首先将模拟得到的气象产量序列按从大到小的顺序排列,依据多元统计分析中对有序样本进行分类的最优分割法^[2],将其划分为3类,分别称为丰年、平年和欠年。相应地,也将历史年份按气象条件划分为气象丰年、气象平年和气象欠年。应用逐步判别分析方法,以气象因子为预报变量,分别建立丰年、平年和欠年3个类别的判别模型,在通过统计检验后,依据气象条件用所得到的判别模型来判断未来年份作物产量的年景趋势。

气象因子初选为4月上旬至9月下旬各旬的气温、降水、日照、积温和由以上4类因子经求和运算构成的组合因子,共计98个。由计算机运行逐步判别分析计算程序从中筛选出进入预报模型的气象因子,并建立预报模型。

2.2 作物产量年景趋势判别模型

1) 玉米产量年景趋势判别模型

丰年模型: $Y = -0190 - 311118x_1 - 2175x_2 + 5104x_3 - 3159x_4 + 3163x_5$

平年模型: $Y = -0197 - 341148x_1 - 3125x_2 + 4196x_3 - 4100x_4 + 4156x_5$

欠年模型: $Y = -1153 - 351110x_1 - 3130x_2 + 5189x_3 - 3165x_4 + 4134x_5$

其中, x_1 为6月下旬平均气温, x_2 为7月上旬到8月上旬积温, x_3 为5月下旬平均气温, x_4 为8月上旬到9月上旬积温。

模型判别能力检验: $U = 0131718$, $V^2 = 28113$ 自由度为 $L(M-1) = 4(3-1) = 8$ 。查表得: $V_{01005}^2(8) = 22$, $V^2 = 28113 > V_{01005}^2(8) = 22$,即由 x_1, x_2, x_3, x_4 建立的判别函数有效。其中, L 为进入模型的变量个数, M 为总体(类别)个数,回代检验37年中有34年预报正确。

2) 大豆产量年景趋势判别模型

丰年模型: $Y = -0197 - 545139x_1 + 0146x_2 + 6102x_3 + 7121x_4 - 13154x_5 + 1165x_6 + 10167x_7$

- 0134x₇+ 0123x₈

平年模型: $Y = - 1129 - 520186 + 0125x_1 + 4141x_2 + 10101x_3 - 12165x_4 + 2134x_5 + 11138x_6 - 0125x_7 + 0120x_8$

欠年模型: $Y = - 1108 - 269148 + 0134x_1 + 5165x_2 + 9131x_3 - 14138x_4 + 1165x_5 + 12133x_6 - 0133x_7 + 0116x_8$

其中, x_1 为 5 月下旬至 6 月下旬降水, x_2 为 6 月上旬平均气温, x_3 为 7 月中旬平均气温, x_4 为 5 月下旬平均气温, x_5 为 7 月下旬到 8 月中旬积温, x_6 为 8 月下旬到 9 月上旬积温, x_7 为 4 月上旬到 5 月上旬降水, x_8 为 8 月降水。

模型判别能力检验: $V^2 = 44176 > V_{0.1005}^2(16) = 3413$, 即由 $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$ 建立的判别函数有效。回代检验 37 年中有 35 年预报正确。

3) 小麦产量年景趋势判别模型

丰年模型: $Y = - 0188 - 3126 + 0110x_1 + 0118x_2 + 0101x_3 + 0106x_4 + 0103x_5 + 0104x_6 + 0110x_7$

平年模型: $Y = - 0197 - 4143 - 0107x_1 + 0120x_2 + 0104x_3 + 0102x_4 + 0106x_5 + 0106x_6 + 0160x_7$

欠年模型: $Y = - 1158 - 7185 + 0110x_1 + 0159x_2 + 0104x_3 + 0102x_4 + 0106x_5 + 0106x_6 + 0160x_7$

其中, x_1 为 4 月份降水, x_2 为 5 月份降水, x_3 为 4 月下旬到 5 月上旬降水, x_4 为 5 月下旬降水, x_5 为 7 月中、下旬降水, x_6 为 4 月上、中旬平均气温, x_7 为 6 月下旬到 7 月下旬积温。

判别能力检验: $V^2 = 44176 > V_{0.1005}^2(14) = 3113$, 判别函数有效。回代检验 37 年中有 35 年预报正确。

4) 实际预报效果检验

运用所得模型对 1995、1996、1997 年 3 年的 3 种作物产量的年景进行预测, 1995 年 3 种作物均为欠年, 与实际情况相符。1996 年为丰年, 也符合实际, 1997 年有一种作物出现误判。预测 1999、2000 两年 3 种作物年景产量情况为: 除 1999 年小麦为欠年, 2000 年为丰年外, 其余两种作物两年均预报为平年。

213 作物产量的定量预报模型

21311 带有周期分量的预报模型^[3]

气象因素的变化, 常常呈现出一定的周期性, 因此, 某些作物的产量也可能具有周期性变化的特点。设有一样本容量为 N 的时间序列 Y_1, Y_2, \dots, Y_N , 按试验周期方法求出试验周期, 其计算过程为: 先将序列 Y_1, Y_2, \dots, Y_N 按周期 2 分成两组, 若 N 为偶数, 则分为 $Y_1, Y_3, Y_5, \dots, Y_{N-1}$; $Y_2, Y_4, Y_6, \dots, Y_N$, 求出每一组的均值 $\bar{Y}_{21}, \bar{Y}_{22}$, 再将序列 Y_1, Y_2, \dots, Y_N 按周期 3 分成 3 组, Y_1, Y_4, \dots, Y_{N-2} ; Y_2, Y_5, \dots, Y_{N-1} ; Y_3, Y_6, \dots, Y_N 求出每组的均值 $\bar{Y}_{31}, \bar{Y}_{32}, \bar{Y}_{33}$, 如此反复下去, 直到分组个数不大于资料序列长度的一半 ($N/2$) 为止, 以各周期的平均值构成的周期序列, 我们定义为周期分量, 设 $M = N/T$ ($N/2$), 则这样的周期分量共计有 $M-1$ 个, 将这 $M-1$ 个序列作为自变量与原序列进行逐步回归运算, 所提取的周期分量称为显著性周期分量, 所得的模型称之为带有周期分量的逐步回归分析模型, 其一般形式为: $Y_i = a_0 + \sum_{i=1}^{M-1} a_i f_i(t) + e_0$ 。3

种作物带有周期分量的产量预报模型

玉米预报模型: $Y = 413 + 016Y_6 + 0150Y_8 + 0130Y_{11} + 0150Y_{12} + 0143Y_{13}, \quad r = 0197, \quad F$

= 84

大豆预报模型: $Y = 1 + 0.160Y_{11} + 0.140Y_{12} + 0.160Y_{13}$, $r = 0.193$, $F = 54$

小麦预报模型: $Y = 0.128 + 0.149Y_{11} + 0.140Y_{12} + 0.147Y_{14}$, $r = 0.196$, $F = 78$, Y_i 表示 i 年周期分量。

运用小麦预报模型对 1996, 1997 两年进行产量预报, 其相对误差在 7% 以下, 说明三江平原地区的小麦产量确存在显著的周期性变化的特点, 这种情况对于其它两个预报点也有相类似的情况。1999, 2000 两年小麦预报产量分别为 2 764 kg Öhm² 和 3 160 kg Öhm²。

2.1.3.12 多层递阶预报模型^[4]

多层递阶预报模型即是以模型

$$Y_k = 5^T(k)H(k) + e_i \tag{1}$$

为基本数学模型。这里参数 $H(k)$ 是随时间 k 而变化的, 称为时变参数。时变参数的确定由下列公式给出

$$H(k) = H(k-1) + \frac{1}{\alpha 5(k)\alpha} 5(k) \{Y_k - 5^T(k)H(k-1)\} \tag{2}$$

其中 $H(k)$ 和 $H(k-1)$ 分别为 $k, k-1$ 时刻的参数值, $\alpha 5(k)\alpha$ 为向量 $5(k)$ 的模。时变参数的算法及多层递阶模型的建立参见文献[4]。

公式(2)的应用首先需要给出初值, 文献[4]已经证明, 在一定的条件下, 预报效果的好坏主要取决于对时变参数的预报精度, 而与(2)式中初值的选择无关。尽管如此, 对于小样本序列来说, 初值的选取仍有可能对预报效果产生影响。

本段预报工作的步骤是: 由时间序列分析中确定模型阶数的 FPE 方法^[4]确定产量预报模型的最佳阶数, 确定参数跟踪公式(2)中的初值 $H(0)$ 并计算出参数序列。对参数序列进行分析并作出预报, 建立产量预报模型。

从多层递阶预报模型的特点可以看出, 模型应用的关键是作出时变参数的预报, 这样, 模型中的参数总是变化的, 每一年的预报模型中的参数都不相同。

1) 玉米产量预报模型

$$Y_k = H(k)Y_{k-1} + H(k)Y_{k-2} + H(k)x_1 + H(k)x_2 + H(k)x_3$$

其中, Y_{k-1}, Y_{k-2} 分别是前一、二年的实际产量, $H(k)$ 是随时间变化的参数, x_1, x_2, x_3 预报年的气象因子, x_1 为 6 月上旬到 7 月中旬的降水, x_2 为 8 月上旬到 9 月上旬的降水, x_3 为 5 月下旬的平均气温。

用此模型预报 1997, 1998 年的玉米产量, 预报值与实际产量值的绝对误差分别为 315% 和 119%。1999, 2000 两年的预报产量分别为 6 146 和 6 237 kg Öhm²。

2) 小麦产量预报模型

$$Y_k = H(k)Y_{k-1} + H(k)Y_{k-2} + HY_{k-3} + H(k)x_1 + H(k)x_2 + H(k)x_3$$

其中, $Y_{k-1}, Y_{k-2}, Y_{k-3}$ 分别是前一、二、三年的实际产量, $H(k), i = 4, 5, 6$ 是随时间变化的参数, x_1, x_2, x_3 为预报年的气象因子, 其含义分别为: x_1 为 6 月下旬到 7 月下旬降水, x_2 为 4 月中旬到 6 月中旬平均气温, x_3 为 7 月中旬平均气温。用所得模型预报 1995, 1996 两年小麦产量, 误差分别为 6% 和 11%。1999, 2000 两年的预报产量为 2 145 和 2 422 kg Öhm²。

3 结 论

本文运用多种统计方法, 对三江平原地区农作物产量的定性、定量预报工作进行了探讨,

得到了粮食产量的定性、定量预报模型, 粮食产量的定性预报模型即逐步判别分析模型, 具有操作简单, 实用性强的特点, 依据气象部门做出下一年度的气象预报, 可快捷方便地得到本年度该类作物的产量年景预报。已在实践中取得较好效果。

粮食产量的定量预报, 既有十分重要的实际意义, 又是难度很大的一项工作。本文尝试使用了多种预报方法, 对于有周期性变化的产量序列, 带有周期分量的预报模型具有一定的优势。多层递阶预报方法, 更好地体现了前期实际的或预报结果, 尤其是将气象因子引入预报模型之中, 使得预报模型不仅含有预报年份前期的产量因子, 同时还含有预报年份的气象因子, 使得预报模型更趋合理。但这种模型的应用比较复杂, 模型的建立也需要大量的计算。使用起来不很方便, 尽管如此, 多层递阶预报方法用在三江平原地区的作物产量预报实践中, 还是有效的。在预报工作实践中, 如果只有产量数据, 应该使用带有周期分量的预报模型, 而在气象资料齐全的情况下, 推荐使用多层递阶预报模型。

[参 考 文 献]

- [1] 农业气象预报文集编审组 农业气象预报文集 北京: 气象出版社, 1983 1~ 32
- [2] 张尧庭, 方开泰 多元统计分析引论 北京: 科学技术出版社, 1982 444~ 457
- [3] 王春乙 用带有周期分量的逐步回归分析模型预测冬小麦产量的试验 气象, 1989(8): 22~ 25
- [4] 韩志刚 多层递阶预报方法及其应用 北京: 科学出版社, 1989 66~ 69, 100~ 203
- [5] 裴鑫德 多元统计分析及其应用 北京: 北京农业大学出版社, 1990 100~ 131
- [6] 罗积玉等 经济统计分析方法及预测 北京: 清华大学出版社, 1987. 181~ 191, 260~ 270

Study on the Forecasting Model of Grain Yield in Sanjiang Plain Region

Liu Zhenzhong Xu Mei

(Heilongjiang August First Land Reclamation University, Mishan, 158308)

Abstract: The cluster analysis is used to zone agroclimatic in Sanjiang plain region, based on the zoning result representative forecasting sites were selected and model for qualitative and quantitative forecast of yields of corn, soybean and wheat were established, including the model of the year's harvest tendency, stepwise regression periodical component model and multilevel recursion forecasting model. Statistical test and forecasting result show that the models can give accurate forecast of the yield the three crops significantly.

Key words: grain yield; yield forecasting model; cluster analysis; the year's harvest tendency