

吉林省农作物产量预测的初步研究*

李明刚 阎日红 梅冬林 陈忠亮

(吉林省农科院综合所)

(吉林省科委)

摘 要

本文运用Fuzzy聚类Bayes判别及逐步回归方法,建立了低中高社会产量阶段模型和丰平欠气象产量预报模式。对全省历年粮食产量进行统计预报表明,拟合准确率达97%以上。

农作物产量预测预报,一直是人们非常关心的课题。准确地预测粮食产量对政府部门指导农业生产,产供销系统科学决策,种植业结构优化等都具有十分重要的意义。在农村实行承包责任制以后,很难准确地统计粮食产量,在这种情况下产量预测预报就更有意义。

作物产量是在人为干预下,作物与其周围环境协调统一的最终表现。它受品种生产潜力,气象条件,土壤因素,社会投入以及栽培管理水平等多种因素的综合影响。因此,仅考虑社会生产条件或气象条件进行产量预报是不够的。本研究旨在综合考虑社会生产因素和气象因素,使产量预测更准确合理,以期为农业生产、科研和政府部门决策服务。

一、资料来源及研究方法

(一) 预报因子的选择

1. 社会产量预报因子的选择 社会产量是指由于社会生产水平逐步提高引起粮食产量稳定上升的部分。为了预报社会产量,在影响社会产量的诸多因素中,经相关和通径分析,筛选出与总产相关密切、直接贡献大的农作物播种面积,玉米播种面积,水稻播种面积,高产作物比例,化肥投入总量和春季保苗率等6个因子,运用1952~1986年统计资料进行社会产量预测。

2. 气象产量预报因子的选择 根据农作物生长发育和产量形成的生物学特性及环境条件的要求,我们选择作物生育关键时期的9项气象因子作为气象产量预报因子。(1) 上年9~11月降水量,关系到第二年春季土壤水分储备。(2) 当年4月中旬至5月下旬气温。(3) 当年4月中旬至5月下旬一次性透雨日数即 $\geq 10\text{mm}$ 日数。这三项因素决定春季农作物幼苗生育状况及保苗率高低。(4) 当年6月中旬至7月上旬降雨量。(5) 当年6月中旬至7月上旬降水日数。这两个因子关系到作物旺盛生长期水热同步状况。(6) 当年8月中旬至9月下旬平均气温。(7) 当年8月中旬至9月下旬降雨量。这两个因子直接影响作物开花授精和子粒形成以及成熟度高低。(8) 当年有效积温。(9) 当年无霜日数。两个因子关系到霜期早晚以及有无延迟性低温冷害发生。运用省气象局历年气象资料建立这九项因子的数据库,再根据因子选择要求,进行计算,作出气象产量预报。

* 本文在李维岳研究员指导下完成,深致谢意!

(二) 研究方法

1. 社会产量预报 根据我省1952~1986年总产变化及社会投入分析,将这34年按生产水平分为低、中、高三个阶段(1952~1964, 1965~1974, 1975~1986年)用逐步回归方法分别建立各阶段产量与社会产量预报因子模型,进行社会产量预报。

2. 气象产量预报 我们把社会产量预报值与历年实际产量之差作为该年气象产量。气象产量预报,首先采用ISODATA Fuzzy分类法,根据前述九项气象预报因子将1952~1986年34年间分为丰、平、欠三类(三类中平均气象产量最大者为丰年,其次为平,最小者为欠年)。然后同逐步回归方法分建立丰、平、欠年气象产量与气象预报因子的预报模式。

3. 丰、平、欠模式识别 采用Bayes方法分别建立丰、平、欠判别函数,将预报因子代入丰、平、欠判别函数,判别预报年属丰、平、欠那一类,然后用相应的模式进行气象产量预报。最后将预报年社会产量预报值和气象产量预报值之和作为预报年总产量预报值。

(三) 数学方法简介

1. 逐步回归(1)由样本值求偏回归系数的增广矩阵,

$$S = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1m} & S_{1y} \\ \vdots & \vdots & & \vdots & \vdots \\ S_{m1} & S_{m2} & \cdots & S_{mm} & S_{my} \end{pmatrix}$$

$$S_{ij} = S_{ji} = \sum_{t=1}^n X_{ti} X_{tj} - \frac{1}{N} \left(\sum_{t=1}^n X_{ti} \right) \left(\sum_{t=1}^n X_{tj} \right)$$

$$S_{iy} = \sum_{t=1}^n X_{ti} Y_t - \frac{1}{N} \left(\sum_{t=1}^n X_{ti} \right) \left(\sum_{t=1}^n Y_t \right)$$

$$(t = 1, 2 \cdots N, i, j = 1, 2 \cdots M)$$

(2)解标准回归系数的正规方程组

$$\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ \vdots & \vdots & & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} r_{1y} \\ \vdots \\ r_{my} \end{pmatrix}$$

$$\text{其中 } r_{ij} = \sqrt{\frac{S_{ij}}{S_{ii}S_{jj}}} \quad r_{iy} = \sqrt{\frac{S_{iy}}{S_{ii} S_{yy}}}$$

(3)计算统计量F反复引入和剔除因子。引入因子时的方差贡献和F统计量:

$$v_i^{(L)} = \frac{[r_{iy}^{(L-1)}]^2}{r_{ii}^{(L-1)}} \quad (\text{对最大者进行F检验})$$

$$F_i^{(L)} = \frac{r_i^{(L)} / 1}{Q^{(L)} / [N(p+1) - 1]} \quad (F_i^{(e)} > F_{\alpha}) \text{则可引入因子}$$

剔除因子时的方差贡献和F统计量:

$$v_i^{(L)} = \frac{[r_{ij}^{(L-1)}]^2}{r_{ij}^{(L-1)}} \quad (\text{对最小者进行F检验})$$

$$F_i^{(L)} = \frac{r_i^{(L)}/1}{Q^{(L)}/(N-P-1)} \quad (F_i^{(e)} < F_{\alpha} \text{ 则剔除因子})$$

$$(4) \text{ 预报方程 } \hat{y} = b_0 + \sum_{k=1}^p b_{ki} X_{ki}$$

$$\text{其中 } b_{ki} = \frac{\sqrt{S_{yy}}}{\sqrt{S_{ki}}} \cdot r_{kiy}^{(L)}$$

$$b_0 = \hat{y} - \sum_{k=1}^p b_{ki} \bar{X}_{ki}$$

2. ISODATA Fuzzy分类 ISODATA Fuzzy是Iterative Self-Organizing Data fuzzy的缩写。

集合x中含有n个样本, 将此n个样本分成C类的分法可用模糊矩阵 U_f 表示

$$U_f = [U_{ij}]_{c \times n}$$

满足: (1) $U_{ij} \in [0, 1]$, U_{ij} 表示 x_i 从属于第i类的从属度

$$(2) \sum_{i=1}^c U_{ij} = 1, \forall j$$

$$(3) \sum_{i=1}^n U_{ij} > 0, \forall i$$

每一个符合(1)、(2)和(3)的矩阵 U_f 对应着一种Fuzzy划分。为了得到最优划分我们希望泛函 $J(U_f, V)$ 为最小, 即聚类准则为:

$$J_m(U_f, V) = \min \{J_m(U_f, V)\} \quad (1)$$

$$J_m(U_f, V) = \sum_{i=1}^n \sum_{j=1}^c (U_{ij})^m \|x_j - v_i\|^2 \quad (2)$$

依Bogdek算法, U_{ij} 和 v_i 的计算判据为

$$U_{ij} = \left(\sum_{k=1}^c \left(\frac{\|x_i - v_i\|}{\|x_j - v_k\|} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (3)$$

$$v_i = \frac{\sum_{j=1}^n (U_{ij})^m x_j}{\sum_{j=1}^n (U_{ij})^m} \quad (4)$$

计算步骤如下:

- 1) 任选一个初始划分 U_f ,
- 2) 用(4)计算 $v_i (i=1, 2, \dots, C)$
- 3) 计算 U_f ,

4) 若 $\max \{ |U_{ij}(t) - U_{ij}(t-1)| \} < \varepsilon (\varepsilon > 0)$ 则 V 和 U 即为所求, 否则回到步骤 2) 再重复同样运算。得到 U_f 以后使 Fuzzy 划分为清晰化的方法是若 U_i , $j = \max(U_{ij})$ 则 x_i 属于第 i 类。

3. Bayes 判别 对于预报对象选好的 P 个预报因子 x_1, x_2, \dots, x_p 。分别对每一类的预报因子进行线性组合, 构成一个线性函数:

$$y = \sum_{k=1}^p C_k X_k$$

C_k 满足下面的方程组

$$S_{11}C_1 + S_{12}C_2 + \dots + S_{1p}C_p = d_1$$

$$S_{21}C_1 + S_{22}C_2 + \dots + S_{2p}C_p = d_2$$

$$\vdots$$

$$S_{p1}C_1 + S_{p2}C_2 + \dots + S_{pp}C_p = d_p$$

$$\text{其中 } S_{k2} = \frac{1}{n_1} \sum_{L=1}^{n_1} [x_{ki}(A) - \bar{x}_k(A)] [x_{Li}(A) - \bar{x}_L(A)] + \frac{1}{n_2}$$

$$\sum_{i=1}^{n_2} [x_{ki}(B) - \bar{x}_k(B)] [x_{Li}(B) - \bar{x}_L(B)]$$

$$d_k = \bar{x}_k(A) - \bar{x}_k(B)$$

设分辨对象为 m 类, 则分别求出其判别函数 $y_k = C_0k + C_{ki}x_i$

$$K = 1, 2, \dots, m, t = 1, 2, \dots, P$$

若 $\hat{y}(k) = \text{Max}(y^{(i)})$ 则预报对象为 K 类。

二、结果与分析

根据算法要求设计逐步回归, ISODATA Fuzzy 分类及 Bayes 判别计算机程序, 用 BASIC 语言, 在 Great Wall 0520 机上运算, 分别得到社会产量和气象产量预报模式。

(一) 社会产量预测模型

1. 低生产水平阶段模型 (1952~1964)

$$y = 125.549 + 1.355x_1 - 2.211x_4 - 0.76x_6$$

$$\text{复相关系数 } R = 0.6753^{**}$$

式中 y 为社会产量 (亿公斤), x_1 为农作物播种面积 (百万亩), x_4 为高产作物比例, x_6 为春季保苗率 (x_4, x_6 取百分数分子部分, 即扩大 100 倍)。此阶段栽培管理水平很低, 社会投入几乎为零, 粮食总产在 42.5~61 亿公斤之间徘徊。

2. 中生产水平阶段模型 (1965~1974)

$$y = 550.973 - 81.818x_1 + 213.335x_2 + 157.343x_3 - 135.036x_4 + 1.348x_5 - 0.609x_6$$

$$\text{复相关系数 } R = 0.8591^{**}$$

式中 x_1 为农作物播种面积 (百万亩), x_2 为玉米播种面积 (百万亩), x_3 为水稻播种面积 (百万亩), x_4 为高产作物比例, x_5 为化肥投入量 (万吨), x_6 为春季保苗率。此

阶段生产水平和社会投入逐步提高,并开始运用优良品种和化肥,粮食总产在49.5~85.5亿公斤范围变动。

3. 高生产水平阶段模型(1975~现在)

$$y = -3483.596 + 54.39x_1 + 4.774x_4 + 0.853x_5 + 0.652x_6$$

$$\text{复相关系数 } R = 0.9097^{**}$$

此阶段生产水平提高很显著,普遍应用高产良种,高产作物比率较高,投肥量从几万吨提高到190万吨,粮食产量提高到72.5~164.5亿公斤

(二) 气象产量预报模型

我们把实际产量与社会产量之离差作为气象条件年际波动引起的产量变化,称为气象产量。气象产量采用综合预报模式。

1. ISODATA模糊分类 为了更准确预报气象产量,我们把前述九项气象因子作为分类指标,由计算机自动给定初始分类矩阵,选取分类精度为0.01,用ISODATA Fuzzy分类方法,将1952~1986这34年分为丰、平、欠三类:

第一类:1952、1967、1970、1975、1981、1982、1983、1984、1985,平均气象产量 = 2.245亿公斤

第二类:1955、1956、1958、1959、1961、1963、1964、1965、1966、1968、1974、1977、1978、1979、1986,平均气象产量 = 1.75亿公斤

第三类:1953、1954、1957、1960、1962、1969、1971、1972、1973、1976和1980,平均气象产量为-5.7345亿公斤

根据各类平均气象产量大小,我们把第一类定为丰年类,第二类定为平年类,第三类为欠年类。

2. 气象产量的丰、平、欠预报模式 分别将丰、平、欠三类中各年分气象产量与九项气象因子资料逐步回归方法建立预报模式。

(1) 丰年气象产量预报模式

$$y = -67.8 - 41.97x_1 + 20.88x_2 + 2.297x_3 + 17.827x_5 + 0.588x_6 - 0.579x_8 + 0.37x_9$$
$$R = 0.9944^{**}$$

其中 y 为气象产量(亿公斤) x_1 为当年4月中旬至5月下旬均温; x_2 为当年4月中旬至5月下旬透雨日数; x_3 为当年6月中旬至7月上旬降水量; x_4 为当年6月中旬至7月上旬降雨日数; x_5 当年8月中旬至9月下旬均温; x_6 为当年8月中旬至9月下旬降雨量; x_7 为上年9月至11月降水量; x_8 为当年无霜日数; x_9 为当年有效积温—MIN。

(2) 平年气象产量预报模式

$$y = 4.009 + 0.805x_1 + 7.275x_2 + 0.564x_3 - 1.383x_4 + 4.062x_5 + 0.165x_6 - 0.204x_7 - 0.553x_8 + 0.049x_9$$

$R = 0.9998^{**}$ $x_1 \sim x_9$ 与丰年预报模式相同。

(3) 欠年气象产量预报模式

$$y = 1243.115 - 19.776x_1 - 27.992x_2 - 1.39x_3 + 6.04x_4 - 30.256x_5 - 3.138x_6 - 0.761x_7 - 2.597x_8 + 0.229x_9$$

$$R = 0.9800^{**}$$

对这三类模型预报结果进行CP(精度)分析表明, 预报效果均达1级。

(三) Bayes法丰、平、欠模式辨识

要进行气象产量预报, 首先必须确定预报年属丰年, 平年还是欠年, 然后用相应的模式预报。为此我们采用Bayes逐级判别方法, 分别建立丰、平、欠年三类判别函数:

$$y_{\text{丰}} = -948.49 + 25.06x_1 + 10.60x_2 + 0.453x_3 - 0.216x_4 + 44.186x_5 + 1.757x_6 + 0.0654x_7 + 3.984x_8 + 0.0035x_9$$

$$y_{\text{平}} = -861.688 + 23.716x_1 + 9.841x_2 + 0.411x_3 - 0.08318x_4 + 44.276x_5 + 1.682x_6 + 0.0514x_7 + 3.714x_8 - 0.03414x_9$$

$$y_{\text{欠}} = -845.072 + 22.714x_1 + 10.290x_2 + 0.426x_3 - 0.4x_4 + 45.4x_5 + 1.689x_6 + 0.0809x_7 + 3.68x_8 - 0.0775x_9$$

$$\text{Wilks统计量 } U = 0.0833$$

$$\text{三级判别卡方 } \chi^2 = 72.064^{**}$$

用Bayes法对1952~1986年34年历史资料丰平欠判别表明, 准确率P=97%以上。

(四) 吉林省历年总产量预报结果

我们用上述社会产量高、中、低阶段模式和气象产量分类辨识丰、平、欠模式对1952~1986年总产进行预报(见表), 表中预测产量=社会产量预测值+气象产量预测值, 产量

历史计算产量与实际产量拟合程度

(单位: 亿公斤)

年代	类别	气象产量	社会产量	预测产量	实际产量	误差	年代	类别	气象产量	社会产量	预测产量	实际产量	误差
1952	1	-4.18	56.57	60.76	61.15	0.40	1970	1	11.25	62.69	73.94	73.90	-0.04
1953	3	0.14	53.53	53.67	56.15	2.48	1971	3	-0.50	71.69	71.44	71.45	-0.01
1954	3	-2.53	55.39	52.86	51.60	-1.26	1972	3	-8.55	63.46	54.91	55.70	0.79
1955	2	2.01	53.56	55.57	55.65	0.08	1973	3	-4.32	83.38	79.02	78.30	0.72
1956	2	-1.01	50.28	49.27	49.35	0.08	1974	2	5.09	80.63	85.73	85.80	0.10
1957	3	-7.00	50.79	43.79	42.95	-0.84	1975	1	-1.53	94.14	92.62	90.65	-1.97
1958	2	6.33	44.44	50.77	52.90	2.13	1976	3	-5.59	83.17	77.58	75.55	-2.03
1959	2	2.19	50.37	52.56	52.65	0.09	1977	2	6.17	66.63	72.79	72.85	0.06
1960	3	-3.54	42.64	39.10	39.45	0.35	1978	2	16.79	72.19	88.98	91.45	2.56
1961	2	-7.04	48.15	41.11	39.85	-1.26	1979	2	2.74	87.65	90.39	90.45	0.06
1962	3	-1.37	44.12	42.45	43.70	1.25	1980	3	-22.65	107.79	85.14	85.95	0.81
1963	2	3.73	46.36	50.09	50.15	0.06	1981	1	-7.63	101.47	93.84	92.20	-1.64
1964	2	0.68	48.45	49.13	49.20	0.07	1982	1	-7.85	106.94	99.10	100.00	0.90
1965	2	-1.36	53.79	52.43	52.50	0.07	1983	1	16.93	130.61	147.54	147.50	-0.06
1966	2	-1.87	61.36	59.69	59.75	0.06	1984	1	16.85	148.82	165.66	164.95	-0.71
1967	1	-0.28	63.95	63.67	64.75	1.08	1985	1	-5.55	127.77	122.22	122.50	0.28
1968	2	5.71	56.43	62.14	62.20	0.06	1986	2	-6.53	145.86	139.33	139.50	0.17
1969	3	-6.69	56.76	50.06	49.85	-0.21							

拟合准确率P=100%

单位为亿公斤。可以看出, 预报准确率和精度都很高。

1987年产量预报结果: 1987年九项气象因子 $x_1 \sim x_9$ 分别为: 12.1, 0, 35.3, 26, 17.5, 51.3, 64.1, 170, 622.8代入判别函数 $Y(1) = 912.14$, $Y(2) = 912.17$, $Y(3) = 898.10$ 因此属第2类, 即1987年为平年。将九项因子代入平年气象产量模型, 气象产量 = 0.341亿公斤, 1987年社会产量预报因子 $x_1 \sim x_9$ 分别为60.533, 31.833, 5.512, 61, 69, 70; 代入高生产水平模型得社会产量 = 158.02亿公斤, 因此, 1987年总产量预报值为158.355亿公斤。

三、讨 论

社会产量高、中、低阶段模式预报和气象产量丰、平、欠辩证模式预报方法, 比较全面地考虑影响农作物产量变化的诸多因素, 能比较准确地反应社会产量与社会生产因素之间、气象产量与关键气象因子之间的定量关系, 不仅用于产量预报, 对于研究粮食产量与制约因素之间的关系方面亦有应用价值。但这种方法只适于预报当年总产量, 有关长期产量预报可采用指数平滑, 残差辨识, 正交多项式以及灰色建模预测等时间序列预测技术。

本研究是“七·五”期间农牧渔业部07课题的一部分, 除本文介绍的方法外, 我们还研制了Cops产量预报系统。该系统提供BDAS数据库管理模块以及多种建模和预测方法。

参 考 文 献

- (1) 王书裕: 《作物产量预报方法》, 1984, 3。
- (2) T·J·Smith etc, statistical-dynamic yield/climate models for Virginia Corn crops, 15th conference on agriculture and forest meteorology, (1981)。
- (3) L·A·Zadeh, "Fuzzy Sets", Information and control 8, 3, (1965)
- (4) 中国科学院大气物理研究所: 《春季连续低温阴雨天气的预报方法》, 科学出版社, 1979。

PRELIMINARY STUDY ON FORECAST OF CROP YIELD IN JILIN PROVINCE

Li Minggang et al.

(Comprehensive Research Institute, Jilin Academy of
Agricultural Sciences)

ABSTRACT

The forecast modes of society yields (high, medium and lower stage) and weather yields (bumper, general and poor harvest) for crops by means of IOSDASA fuzzy classification, Bayes distinguishing method and step-wise regression were established and reported in this paper. The result showed that the similar percentage of simulant forecast was more than 97% based the statistic of crop yields over the years and the correlation analysis of yields with social and weather elements.