

# 河北省穗期麦蚜预测及种群动态模拟

郭金霞 李建成 赵文臣\* 李智惠<sup>1</sup> 谷兆昱<sup>1</sup> 石少亭<sup>2</sup> 孔祥兰<sup>2</sup>

(河北省农林科学院植物保护研究所,保定 071000)

**摘要** 采用聚类代表元法筛选自变量,根据逐步判别分析和逐步回归原理,通过计算机模拟技术,对河北省麦蚜主发生区田间系统调查资料进行了统计分析,组建了麦蚜发生程度4级判别式,应用该式可提前2个月预测穗期蚜量。经对保定、正定的历史数据回判,准确率达100%。对麦蚜种群数量增长研究结果表明,5月份百茎蚜量随时间变化呈“S”型,可用 Logistic 曲线方程表述。本研究结果为麦蚜综合防治提供了决策依据。

**关键词** 麦蚜 聚类代表元 统计分析 计算机模拟 预测模型

进入80年代后,我国麦蚜发生趋向严重,一般造成小麦减产20%,重者达30%,成了生产上的一大问题。前人对麦蚜预报进行了研究,但多是根据经验短期预测,准确率低,难以指导防治<sup>[1]</sup>。为此,我们在弄清麦蚜发生规律、种群动态及为害程度的基础上<sup>[2]</sup>,选择河北省具有代表性的正定、保定、丰南等地麦田作为调查点,依据多年研究资料,采用聚类代表元法<sup>[3]</sup>,从36个因子中选出若干个所含元素较多且与峰值蚜量有较大相关的“代表元”作为预报因子,以期克服仅根据农事方面的经验来选取变量的局限性,提高预测的准确性和测报水平。同时,为长期因子的选择探索新的途径。

## 1 研究方法

### 1.1 数据采集

在田间无人工干扰条件下,于保定、正定、丰南分别设麦蚜系统调查圃。从麦蚜始见期开始,每5d调查一次,每次调查田间5点取样,每点20茎。共计调查100茎,记载蚜量,直到小麦成熟为止。

根据多年调查资料,以 $x$ 表示小麦灌浆期百茎高峰蚜量(单位,头/百茎),将麦蚜发生程度分为4个级别:轻(1级): $x \leq 1000$ ;中(2级): $1001 \leq x \leq 2000$ ;重(3级): $2001 \leq x \leq 3500$ ;大(4级): $x \geq 3501$ 。

根据现有资料,每个级别都有4~6个重复,这样获取的数据具有代表性。

### 1.2 变量筛选

设 $y_1, y_2, \dots, y_{21}$ 依次为保定1980~1990年、正定1981~1990年5月份百茎最高蚜量; $x_1, x_2, \dots, x_{12}$ 分别代表上年7月至当年6月月平均气温; $x_{13}, x_{14}, \dots, x_{24}$ 分别表示上年7月至当年6月月平均相对湿度; $x_{25}, x_{26}, \dots, x_{36}$ 分别表示上年7月至当年6月月平均降水量(即

月总降水量的  $\frac{1}{3}$ 。

为了解决从 36 个因子中筛选变量的问题, 我们采用了“聚类代表元”法, 即首先计算 37 阶相关矩阵  $R$ , 将每一行不小于  $r$  的  $x_i$  列出, 与  $x_j$  组成一类记作  $C(i, k_j)$ , 根据  $k_j$  的多少及  $x_i$  与  $y$  的相关系数进行变量剔除。

### 1.3 组建预测模型

根据逐步判别分析, 逐步回归原理<sup>[4]</sup>, 建立麦蚜发生程度判别式, 高峰蚜量预测式。在此基础上应用 Logistic 曲线模拟 5 月份蚜量随时间变化的消长曲线。

### 1.4 计算机模拟

(1) 菜单提示; (2) 数据处理; (3) 计算结果。

## 2 结果与分析

### 2.1 麦蚜发生程度判别

将 36 个因子 21 次重复的数据输入“因子筛选”程序。取  $\alpha=0.9$ , 因  $n=21$ , 自由度=19, 查得相应  $r=0.3687$ , 按照聚类代表元法规则剔除因子后, 得到了以下 13 类代表元:

$C(1): (x_{32}   x_{18}, x_{19}, x_{20}, x_{21}, x_{30}, x_{33}, x_{34})$	$r_{32}=0.265$
$C(2): (x_{20}   x_{18}, x_{19}, x_{21}, x_{30}, x_{33}, x_{34})$	$r_{20}=0.363$
$C(3): (x_{13}   x_7, x_8, x_{11}, x_{14}, x_{25}, x_{30})$	$r_{13}=0.271$
$C(5): (x_{27}   x_7, x_{15}, x_{16}, x_{25}, x_{28})$	$r_{27}=0.046$
$C(6): (x_{18}   x_{19}, x_{21}, x_{24}, x_{30}, x_{33})$	$r_{18}=0.527$
$C(7): (x_8   x_7, x_{10}, x_{11}, x_{19}, x_{31})$	$r_8=0.527$
$C(8): (x_4   x_2, x_7, x_{23}, x_{25})$	$r_4=-0.253$
$C(10): (x_{28}   x_5, x_{15}, x_{16})$	$r_{28}=-0.286$
$C(11): (x_{26}   x_{10}, x_{11}, x_{14})$	$r_{26}=-0.142$
$C(13): (x_{17}   x_2, x_{12}, x_{29})$	$r_{17}=-0.347$
$C(17): (x_{30}   x_{21})$	$r_{30}=0.566$
$C(19): (x_{16}   )$	$r_{16}=-0.553$
$C(34): (x_5   )$	$r_5=-0.584$

将以上 13 个代表元作为相关因子, 代入逐步判别程序, 计算机显示如下结果:

$$\begin{aligned}
 y_1 &= -2532.46 + 13.9221x_{18} + 43.9935x_{13} - 12.2099x_{27} + 13.916x_{17} \\
 &\quad - 223.896x_{30} - 26.835x_8 - 67.5049x_4 + 17.2069x_{16} \\
 y_2 &= -2316.54 + 13.6246x_{18} + 41.8362x_{13} - 11.5078x_{27} + 13.4346x_{17} \\
 &\quad - 213.667x_{30} - 26.0544x_8 - 64.3513x_4 + 16.2433x_{16} \\
 y_3 &= -2474.44 + 14.1966x_{18} + 43.7674x_{13} - 12.004x_{27} + 13.7902x_{17} \\
 &\quad - 224.11x_{30} - 27.8984x_8 - 68.555x_4 + 16.6619x_{16} \\
 y_4 &= -2237 + 13.2164x_{18} + 41.5581x_{13} - 11.3057x_{27} + 13.0972x_{17} \\
 &\quad - 208.946x_{30} - 25.0156x_8 - 64.1202x_4 + 15.8936x_{16}
 \end{aligned}$$

式中  $y_1(\cdot), y_2(\cdot), y_3(\cdot), y_4(\cdot)$  是判别式, 设  $\dot{x} = (\dot{x}_{18}, \dot{x}_{13}, \dot{x}_{27}, \dot{x}_{17}, \dot{x}_{30}, \dot{x}_8, \dot{x}_4, \dot{x}_{16})$  是待判数据, 则  $\dot{x}$  类别所属的判别方法是: 若  $y_i(\dot{x}) > y_j(\dot{x}), i=1, 2, 3, 4$ , 则  $\dot{x}$  判属之  $i$  类。数学上能保证  $y_{(i)}(\dot{x}) = y_{(j)}(\dot{x})$  的概率为 0, 故一般情况下总可判别出来。

用上述判别式对历史数据进行回判, 准确率为 100% (表 1)。

表1 1980~1990年预测与实测结果

地点		1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
保定	实际值	4	1	4	1	3	2	4	3	2	4	3
	预测值	4	1	4	1	3	2	4	3	2	4	3
正定	实际值	—	1	3	1	1	2	2	4	3	1	3
	预测值	—	1	3	1	1	2	2	4	3	1	3

## 2.2 峰值蚜量预测

麦蚜发生程度判别,对麦蚜宏观管理,综合防治是必要的,但对各地具体情况来说,仅有这一点是不够的。5月份高峰蚜量是麦蚜发生轻重的重要指标变量。根据正定县1981~1990年田间麦蚜发生情况的系统调查资料,对穗期峰值蚜量预测模拟,仍以36个因子作为变量筛选对象,经过逐步回归分析后,得到如下预测模型:

$$y = 7942.06 + 140.994x_9 + 19.6834x_{19} - 472.036x_4 + 188.645x_{30} - 348.958x_5 \\ + 86.4369x_{29} \quad s = 89.3 \quad r = 0.999$$

在实际判别中,若残差(实际值与回归值之差)小于 $3s$ ,即为判别正确。

根据预测模型对正定麦蚜历史发生情况进行回判(表2)。

表2 麦蚜发生数量的回判结果 (正定)

	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
回归值	192.36	2371.00	676.33	268.88	1789.00	1964.77	4608.19	2478.14	312.38	2864.84
残差	-94.86	-34.00	5.66	74.12	45.00	-17.77	39.81	-9.74	41.61	-49.84

## 2.3 预测实例

应用以上结果,我们对正定、保定两地1992年麦蚜发生情况进行了预测,有关气象资料见表3。

表3 1991~1992年与预测有关的气象资料

年度	月份	月平均相对湿度(%)		月平均气温(℃)		月降水量(mm)	
		正定	保定	正定	保定	正定	保定
1991	7	$x_{13}$	75	77			
	9					$x_{27}$	48.1
	10	$x_{16}$	54	58	$x_4$	14.6	13.8
	11	$x_{17}$	56	58	$x_5$	5.3	4.8
	12	$x_{18}$	74	71		$x_{30}$	8.3
1992	1	$x_{19}$	58				
	2			$x_8$	2.5	2.1	

将表中 $x_{18}$ 、 $x_{13}$ 、 $x_{27}$ 、 $x_{30}$ 、 $x_8$ 、 $x_4$ 、 $x_{16}$ 的值依次代入判别式,判别结果为正定、保定两地1992年麦蚜均为4级发生程度。实际发生情况是两地峰值蚜量分别为4046头/百茎和5400头/百茎。均大于3501头/百茎,为4级发生,判别正确。

将正定 $x_8$ 、 $x_{19}$ 、 $x_4$ 、 $x_{30}$ 、 $x_5$ 、 $x_{29}$ 代入蚜量预测方程,得到预测值为 $5408.95 \pm 267.9$ 头/百茎。

从以上结果看,无论是对历史发生情况的回判,还是对 92 年的预测模拟,预测值与实际值的吻合很好。只有 92 年发生数量预测值偏高但仍在同一级别之内,说明该式具有较高的实用价值。

#### 2.4 种群数量动态模拟

根据多年的观察资料,我们发现不同年份的麦蚜发生程度,其种群数量增长是有差异的。为了准确地模拟 5 月份麦蚜种群数量随时间变化的曲线,我们在麦蚜发生程度判别的基础上进行分级模拟,根据丰南 1985~1990 调查资料,发现该地 6 年间有 4 年都是大发生年,即 5 月份百茎最高蚜量大于 3500 头。只有 1985、1988 两年属中等发生程度,并且发现大发生年份麦蚜种群数量增长曲线呈“S”型,可用 Logistic 方程进行拟合。

将丰南麦蚜大发生年 5 月份麦蚜种群数量与时间变化关系数据输入计算机“Logistic”程序运行后得到如下结果:

$$y = \frac{7764.519}{1 + 2843.669e^{-1.32027T}} \quad \text{均方误差} = 15.45967$$

式中,  $y$  为百茎蚜量,  $T=1$  即为 4 月 30 日,以后每 5 天  $T$  值增加一个单位。

拟合效果见表 4。

表 4 5 月份麦蚜种群数量与时间变化关系式的拟合效果

T	实际值	拟合值	残 差
1	6.00	10.21	-4.21
2	27.00	38.09	-11.90
3	113.00	140.72	-27.72
4	524.00	501.92	22.08
5	1590.00	1596.14	-6.14
6	3821.00	3820.83	0.17

### 3 讨论

对于农业害虫种群动态消长的预测预报,应该是长、短期预报相结合,我们应用“聚类代表元”法比较成功地解决了因子筛选问题,能在虫害发生当年 2 月份对麦蚜发生程度进行轻、中、重、大四级判别,对 5 月份高峰蚜量进行预测,经过多年历史数据进行回判准确率为 100%,具有较强的实用价值。

蚜量随时间变化的曲线对于依据防治指标确定具体防治时期是非常必要的,在发生程度早期判别的基础上,用 Logistic 曲线对大发生年份进行了模拟,预测值均方误差较小,拟合曲线应用效果较好,对于其它级别的蚜量与时间的变化曲线模拟尚需有多年资料进一步研究。

“聚类代表元法”在预测因子的筛选中是一种比较合理的方法,因为它在筛选代表元时是从两方面考虑的,一是根据所需  $\alpha$ ,使代表元尽量多地与其它元素  $\alpha$ —线性相关;二是使代表元与峰值蚜量的相关系数相对大些。所以,这样选出的变量信息量大。它不仅反映了变量本身与峰值蚜量的直接函数关系,而且蕴藏着比较多的变量信息,使模型的拟合准确率高,从而提高了预报水平,能科学地指导防治。

## 参 考 文 献

- 1 张树波等. 镇原县麦蚜发生规律及预测模式. 甘肃农业科学, 1991(2): 20
- 2 赵文臣等. 小麦长管蚜消长动态及经济阈值. 华北农学报(增刊), 1989: 140~143
- 3 Ping Sun. Classified representative method—a useful way to select variables in ecology. *Ecological Modelling*, 1991(55): 123~130
- 4 唐守正. 多元统计分析法. 北京: 中国林业出版社, 1986: 164~182, 259~263

## Aphid Population Dynamic and Long-term Prediction in the Wheat Winter Field in Hebei Province

Guo Jinxia   Li Jiancheng   Zhao Wenchen  
Li Zhihui   Gu Zhaoyu   Shi Shaoting   Kong Xianglan

(*Institute of Plant Protection, Hebei Academy of  
Agricultural and Forestry Sciences, Baoding*)

**Abstract** The field census data of aphid population in Zhengding county and Baoding area, Hebei province, was statistically analysed by use of the principle of stepwise regression and stepwise discrimination and the method of representative elements of clustering analysis for selecting the independent variables, and the four-grade discrimination formula for reporting the aphid disaster degree and the long-term prediction models for reporting the highest aphid density in May were established. The aphid population at the date of heading could be predicted two months ahead of time. In order to determine the reliable control date, the growth regularity of aphid population in May was studied. The results showed that the aphid population growth curve assumed a "S" type, and the logistic curve equation was  $Y = \frac{7764.519}{1 + 2843.669e^{-1.32027x}}$ . The above results can provide basis for controlling aphid in winter wheat field.

**Key words:** Wheat aphid; Representative elements for clustering analysis; Statistical analysis; Computer simulation; Prediction model