

华北地区冬小麦产量 预报模式的初步研究*

宛公展 王宝成 蓝鸿第

(天津市气象局) (天津市) (国家气象局)
气象科学研究所 塘沽气象台 气象科学研究院

提 要

本文提出了一种省级以上尺度的作物-天气产量预报模式。首先,对气候产量进行 R-型聚类分析,将全区分成若干亚区;然后,用累加型及阶乘型的多重非线性回归方法组建预报方程。本文提出的模式,原则上可以适用于多种尺度和不同作物,预报时效也可以根据生产需要和作物发育阶段的特点给定。

华北地区冬小麦产量预报的时效,分为冬前、越冬后及黄熟前三个阶段。预报方程的复相关系数,分别可达到 0.70, 0.90 和 0.95 以上;平均相对剩余误差为 15%, 8% 和 4%。

一、引 言

农业气象预报,是农业气象工作中经济效果最明显的分支之一。近二十年来,我国逐步开展了作物播种期、发育期以及作物产量预报等工作,有些已进行试报供有关部门参考,并取得良好效果^[1-4]。冬小麦是我国主要粮食作物之一,华北地区是冬小麦主要产地,但产量很不稳定。本文用华北地区 42 个县(市)1955—1979 年实际单产与同期气象资料,建立了冬小麦-天气产量预报模式。

二、时间趋势产量的确定

某种作物的产量,是当地的土壤肥力、农艺措施以及气象因子等综合作用的结果。一般讲,随着农业生产水平的提高,产量有逐年上升的趋势,同时还叠加着某些随机的波动。这些波动,在很大程度上可以认为是气象因子的影响。因此,作物的实产 Y 可以分解为时间趋势产量 Y_t 和气候产量 Y_w 以及随机误差 e , 即

$$Y = Y_t + Y_w + e \quad (1)$$

* 本文于 1982 年 6 月 16 日收到, 1983 年 8 月 15 日收到修改稿。

产量的时间趋势项 Y_t , 通常多采用正交多项式模拟^[5-6]。但该方法农业意义不很清楚, 按时间外延时所担风险也较大。尽管 Y_t 有逐年升高的趋势, 但不会无止境地增大下去。随着逐年生产水平及农艺措施的不提高与改进, 伴有一个逐渐增强的负反馈作用于 Y_t , 使得 Y_t 渐渐趋于一个常数 Y_c , 这就是当地可以预期的最高产量。这个过程, 可以用下面微分方程来模拟, 即

$$\frac{dY_t}{dt} = kY_t(Y_c - Y_t) \quad (2)$$

其中, k 为比例系数。(2) 式的解为

$$Y_t = \frac{Y_c}{1 + \exp[A + Bt]} \quad (3)$$

(3) 式, 即为 Logistic 函数^[7]。 Y_c, A, B 为待定参数。这是一条 S 型曲线, 当 t 较小时, Y_t 增长较慢; 当 t 增大到一定数值后, Y_t 增长最速; 但 t 再增大, Y_t 的增长又减缓下来; 特别当 $t \rightarrow \infty$, $Y_t \rightarrow Y_c$ 。这与大多数地区作物单产时间序列的变化趋势颇为一致。即解放初期, 由于生产力水平较低, 单产增长较慢; 以后随着生产力的发展, 单产增长较为迅速; 但当产量增加到一定水平以后, 再大幅度地提高单产, 很不容易。因此, 单产的增长又缓慢下来, 并趋近于 Y_c 。我们认为, 用这条曲线来描述产量的时间趋势, 似较合理。外延时, 所担风险较小。为便于计算, (3) 式中的 Y_c , 我们用历史最高产量, 加上按直线回归拟合实产时的一个剩余标准差 δ 来确定, 即

$$Y_c = Y_{\max} + \delta \quad (4)$$

表 1 华北地区部分台站时间趋势产量的主要参数

台 站	参 数	Y	A	B
天 津		350.3	1.506	-0.090
保 定		627.3	1.948	-0.137
石 家 庄		814.1	1.719	-0.142
太 原		352.9	1.003	-0.080
安 阳		625.1	1.314	-0.108
卢 氏		230.7	0.636	-0.081
许 昌		765.5	1.841	-0.122
南 阳		510.2	1.529	-0.109
信 阳		528.0	1.393	-0.067
济 南		474.6	2.045	-0.123
泰 安		569.8	1.976	-0.124
昌 潍		970.9	2.531	-0.153
莱 阳		455.5	1.873	-0.100
烟 台		840.9	1.547	-0.110
临 沂		399.3	1.469	-0.082
青 岛		478.1	2.269	-0.150
沈 阳		225.7	1.299	-0.086
大 连		416.7	1.100	-0.046

A 、 B 值的确定,可对实产资料 Y 用 Logistic 函数进行最小二乘拟合即可。表 1 中,列出了华北地区部分台站的 Y_c 和 A 、 B 值。

三、气候产量的提取和聚类

由(1)式,单站的气候产量估值 $\hat{Y}_w$,可写成

$$\hat{Y}_w = Y - Y_i \quad (5)$$

按(5)式提取的气候产量有正、有负。为方便以后的数据处理,可将 $\hat{Y}_w$ 都变为正值。这时,只需将 Y_i 沿纵坐标轴,下移两个剩余标准差即可基本解决。

由于华北地区空间尺度较大,冬小麦生长、发育和形成产量的农业气候条件不尽相同。因此,即使同一年,不同的地方可能彼丰此欠。既然地区间农业气候条件的不同,在很大程度上可以通过 $\hat{Y}_w$ 的差异反映出来。因此,我们首先对 $\hat{Y}_w$ 进行分类,将 $\hat{Y}_w$ 相近似的台站划为同一亚区,再按亚区分别组建预报方程。这样,即可避免逐站处理资料时面临的大量繁杂计算,又可以增加回归方程的相对稳定性。

对 $\hat{Y}_w$ 进行客观地分类,我们采用 R-型聚类分析的方法^[8]。比较变量之间相似的程度,常用的距离指标有相关距离与欧氏距离等。相关距离是两个向量相关系数的反余弦,欧氏距离为两向量各分量差值的平方和。它们分别侧重于考虑变量间的比例关系与绝对差异,各有一定的局限性。为兼顾二者之优,我们提出以下的综合距离指标

$$d_{ij} = (1.5 - 0.5 r_{ij}) \sqrt{\frac{l_{ij}}{n}} \quad (6)$$

$$(i, j = 1, 2, \dots, n)$$

这里, n 为台站个数, r_{ij} 与 l_{ij} 分别表示台站 i 与 j 之间的单相关系数与欧氏距离。由元素 d_{ij} 组成的距离矩阵 D 为 n 阶对称方阵,对角线元素均为 0。

进行聚类分析时,首先在 D 阵中(不包括对角线元素)挑取一个极小值 d_{pq} ,不失一般性, $p < q$, 此时 p, q 两类合并,称为新 p 类(高类一律向低类合并)。然后,用类平均法,计算并类后的 p 类(即新 p 类)与其余各类的距离,变换公式为

$$\begin{cases} A = \frac{K_p^{(l-1)}}{K_p^{(l-1)} + K_q^{(l-1)}}, B = \frac{K_q^{(l-1)}}{K_p^{(l-1)} + K_q^{(l-1)}} \\ d_{ip}^{(l)} = A \cdot d_{ip}^{(l-1)} + B d_{iq}^{(l-1)} & (i = 1, 2, \dots, p-1) \\ d_{pj}^{(l)} = \begin{cases} A d_{pi}^{(l-1)} + B d_{qj}^{(l-1)} & (\text{当 } q < j) \\ A d_{pj}^{(l-1)} + B d_{iq}^{(l-1)} & (\text{当 } q > j) \end{cases} \end{cases} \quad (7)$$

$$(j = p+1, p+2, \dots, n; j \neq q)$$

(7)式中, K_p, K_q 代表各类台站权重矩阵中 p, q 两类的权重值。初始时, $K_j = 1 (j = 1, 2, \dots, n)$ 。随着聚类的进行,权重矩阵 K 不断变化。比如聚类计算至第 l 步、 q 类合并至 p 类。此时, $K_q^{(l)} = 0, K_p^{(l)} = K_p^{(l-1)} + K_q^{(l-1)}$; 这里的 $(l-1)$ 表示 l 步的前一步。该算法的每一步计算,只需对 D 阵上三角区域的 p 行与 p 列元素,实施(7)式的变换即可。当重复计算至 $n-2$ 步时,全部台站聚为两大类,聚类结束。

用(6)、(7)式,我们对华北地区各县(市)气候产量进行聚类计算,其结果可以绘成聚

表 2 华北地区冬小麦气候产量的分区

聚 类 分 区	主 要 台 站
一 亚 区	洛阳、新乡、卢氏、商丘、许昌、南阳、驻马店、信阳、运城
二 亚 区	烟台、莱阳、昌潍、青岛、临沂、菏泽、济宁、泰安、济南、德州、临清、 安阳、滨县、
三 亚 区	唐山、天津、沧县、衡水、邢台、邯郸、长治、介休、离石、太原、忻县、 河曲、大同、张家口、保定、石家庄
四 亚 区	大连、沈阳、营口、锦州

类枝型图(图略)。根据此图,同时参考自然地理及行政区划等,我们将华北冬小麦产区分四个亚区,详见表 2。然后,每个亚区按同样的步骤分别建立适用于本区的预报方程。

四、多重非线性回归模式的建立

投入业务使用的产量预报模式,要求能在作物不同的发育阶段,提供精度不同的潜在产量估计。而且,随着预报时效的缩短,精度还应逐步提高。这集中体现在,能不断地用现在有利或不利的天气条件去订正前一时期的预期产量,并使得不同时期的预报值逐步逼近真值。这样,一般的多元回归或积分回归模式已不能胜任。为此,我们考虑可以试用以下多重非线性回归模式,其基本模型有两种,即

$$\hat{Y}_{w_t} = V_0(\hat{Y}_{w_{t-1}}) + \sum_{i=1}^p V_i(X_i) \quad (8)$$

$$\hat{Y}_{w_t} = V_0(y_{w_{t-1}}) \prod_{i=1}^p V_i(X_i) \quad (9)$$

(8)式称为累加型(或 0 型), (9)式为阶乘型(或 1 型)。 $\hat{Y}_{w_{t-1}}$ 为前一发育阶段的预期产量。从(8)、(9)式不难看出, t 时刻的预测产量,是用已出现的气象因子 X_i , 对前一时刻($t-1$)的估产值 $Y_{w_{t-1}}$ 用差值法或比值法进行多重订正。至于 $\hat{Y}_{w_{t-1}}$ 的初始值(即第一次作预报),可分别取 0 和 1。(8)与(9)中的函数形式 $V_i(X_i)$, ($i=1, 2, \dots, p$), 一般可取抛物线形式,这是基于作物气候生态三基点的假设。即作物生长、发育和形成产量的农业气候条件,有下限、最适和上限之别。它们可以近似地用抛物线来表征^[9-12], 即

$$V_i(X_i) = b_{i0} + b_{i1}X_i + b_{i2}X_i^2 \quad (10)$$

除规定上一阶段的预期产量 $Y_{w_{t-1}}$ 作为预报因子最先进入方程外,其余气象因子 X_i , 都是按重复回归的方法依次进入方程的。其次序,是按照它们与气候产量的相关指数(按二次三项式曲线回归计算)大小排队确定。凡相关指数较大的因子,优先进入方程。这里顺便指出,作多重回归计算的气候产量与气象因子 X_i , 均为某亚区的台站平均值。

为使得进入方程的因子尽量相互独立,每引入一个新的因子之前,须先鉴定一下它与已经进入方程的因子间的相关程度,凡相关系数大于临界值 r_c (根据样本大小确定)时,表明该因子独立性较差,将其筛掉。否则,才能进入方程,并计算出相应的回归系数 b_{ij} ($j=$

0, 1, 2) 值。

为了尽可能利用气象因子提供的信息量, 我们还对初始因子进行膨化处理。比如冬小麦某生育阶段的月平均气温顺序为 $T_1, T_2, \dots, T_q$; 显然, 除了每月温度各为一个因子共 q 个因子外, $T_1 + T_2, T_2 + T_3, \dots, T_{q-1} + T_q; T_1 + T_2 + T_3, T_2 + T_3 + T_4, \dots, T_{q-2} + T_{q-1} + T_q; \dots, T_1 + T_2 + \dots + T_q$; 又组成了各具不同农业意义的新因子, 它们可以分别表示某阶段或全生育时段的累积温度。因此, 经膨化处理后, 原来的 q 个因子将变为 $q(q-1)/2$ 个。另外, 为使得(8)、(9)有较高的稳定性, 我们还控制进入方程的因子总数, 一般不超过样本容量的 $1/5$ 。

五、计算结果与试报

将华北地区冬小麦全生育期分为 11 个时段(从当年 7 月至翌年 5 月, 每月为一段), 气象因子选取月平均气温 T (9—5 月), 月降水量 R (7—5 月) 及月日照时数 S (9—5 月) 共三类。预报发布时间为 11 月初(先考虑播种条件)、3 月初(再考虑越冬情况)和 6 月初(最后考虑穗分化及籽粒灌浆等)共三次。产量与气象资料均取 1955—1979 年, 并取出 1979 年样本不参加统计分析, 留做试报。

现以一亚区为例, 将经过多重回归分析得到的预报方程及主要参数, 列在表 3。其余各亚区类同, 不再赘述。从表 3 可以看出, 11 月初发布第一次产量预报时, 选中的因子有 R_9 (9 月降水量), R_7 (7 月降水量), R_{8+9} (8 月 9 月降水总和) 和 T_{10} (10 月平均气温)。这不是偶然的, 如所周知, 冬小麦冬前形成壮苗, 可为丰收奠定基础。而壮苗的形成, 与底墒

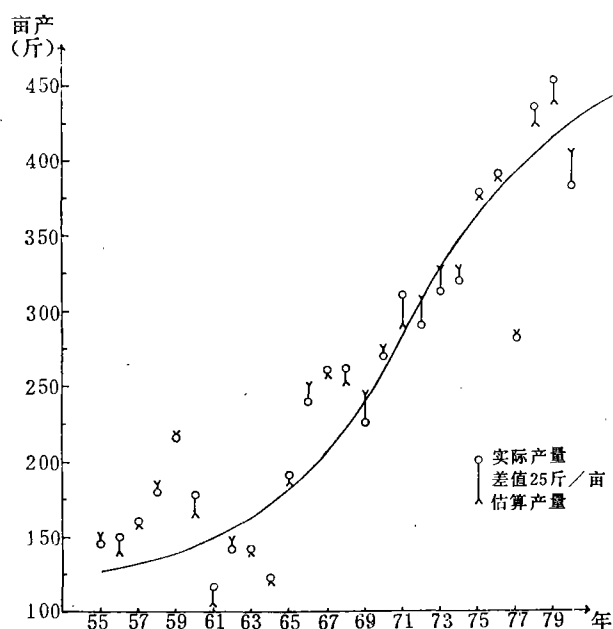


图 1 华北地区一亚区实际产量与黄熟期用 1 型方程估算的产量
以及产量的时间趋势(用 Logistic 函数拟合)

表 3 华北地区一亚区冬小麦产量预报方程的主要参数

类型	时 效	选中因子	回 归 系 数			复相关系数 R	剩余标准差	
			b_0	b_1	b_2		(斤/亩)	(%)
累 加 型 (0)	11 月初 (播种后)	R_0	169.30	-1.29	0.0052	0.712	30.7	12.2
		R_7	48.19	-0.67	0.0019			
		R_{8+9}	-25.46	0.24	-0.0004			
		T_{10}	1691.50	-220.47	7.1532			
	3 月初 (越冬后)	$\hat{Y}_{w,-1}$	-69.85	2.09	-0.0039	0.935	15.5	6.1
		R_{12+1+2}	-53.01	2.10	-0.0170			
		T_2	2.31	6.57	-1.7619			
		T_{1+2}	-3.89	-0.42	0.3551			
	6 月初 (黄熟期)	S_{11}	109.57	-1.45	0.0046	0.976	9.5	3.8
		$\hat{Y}_{w,-1}$	16.08	0.63	0.0017			
		S_{3+4+5}	49.95	-0.14	0.0000			
		T_4	-143.97	20.98	-0.7559			
阶 乘 型 (1)	11 月初 (播种后)	R_0	169.30	-1.29	0.0052	0.661	32.9	13.0
		R_7	1.42	-0.00	0.0000			
		R_{8+9}	0.72	0.00	-0.0000			
		T_{10}	16.20	-1.97	0.0636			
	3 月初 (越冬后)	$\hat{Y}_{w,-1}$	-29.17	1.43	-0.0013	0.924	16.7	6.6
		R_{12+1+2}	0.37	0.02	-0.0001			
		T_2	1.05	0.03	-0.0129			
		T_{1+2}	0.95	-0.00	0.0044			
	6 月初 (黄熟期)	S_{11}	1.58	-0.00	0.0000	0.976	9.5	3.8
		$\hat{Y}_{w,-1}$	-4.81	1.07	-0.0002			
		S_{3+4+5}	3.32	-0.00	0.0000			
		T_4	-2.08	0.42	-0.0144			
		T_5	3.37	-0.20	0.0043			
		R_3	1.03	0.00	-0.0000			

表 4 1979 年各亚区不同时效的试报结果

单位: 斤/亩

亚 区	实 产	类 型 预 报 时 效	0 型 方 程			1 型 方 程		
			11 月初	3 月初	6 月初	11 月初	3 月初	6 月初
一	451.4	预报值	391.0	424.6	431.4	388.0	426.7	437.9
		差 值	-60.4	-26.8	-20.0	-63.4	-24.7	-13.5
二	511.6	预报值	438.5	479.6	485.8	438.4	475.9	489.2
		差 值	-73.1	-32.0	-25.8	-73.2	-35.7	-22.4
三	481.3	预报值	413.9	465.1	472.4	408.3	463.0	484.4
		差 值	-67.4	-16.2	-8.9	-73.0	-18.3	-3.1
四	147.5	预报值	172.7	150.9	148.0	167.4	149.1	148.6
		差 值	25.2	3.4	0.5	19.9	1.6	1.1

(取决于伏天7—8月的降水量),播种时的墒情(9月降水量)及冬前的温度(10月温度)状况密切相关。3月初发布二次预报时,选中因子有 R_{12+1+2} (冬季雨雪总量), T_{1+2} (冬季1月2月气温之和)等。这些因子与冬小麦越冬条件的优劣相关密切。6月初最后一次预报时,选中因子主要为 S_{3+4+5} (春季日照时数总和), T_4 (4月气温), T_5 (5月气温)和 R_3 (3月降水量)等。这些因子对冬小麦幼穗分化及籽粒灌浆等都有一定的影响。由表3还可以看出,无论0型还是1型的回归方程,随预报时效的缩短,都可以期望预报精度不断改善,剩余误差逐渐降低,计算出气候产量再加上时间趋势产量,即可得到实产的估值。本亚区用1型方程在黄熟期进行的估产与实际产量的拟合情况,详见图1。另外,我们还分别统计了本亚区0型与1型方程在黄熟期发布预报的拟合误差,其中误差小于10斤/亩的,均约各占80%。这样看来,0型与1型模式,最终预报精度没有显著差异,只是在预报时效较长时,0型略优于1型。

1979年试报结果,各亚区均列在表4。从该表可以看到,不同时效的预报误差,分别为20—70斤/亩,2—32斤/亩和1—26斤/亩左右。

六、结 语

尺度较大地区的作物产量预报,首先必须解决点与面的关系。用聚类分析的方法将大地区较客观地分成若干亚区(每亚区可以包括一个或数个县),我们认为较好的途径之一。由于各亚区之间的农业气候条件有较显著的差异,因此各成相对独立的预报体系。

用多重非线性回归的模式,解决不同时效的产量估算及其互相衔接,方程的农业意义清楚,预报效果较佳。特别是,由于时间趋势产量 Y_t 是逐县(市)计算的,而气候产量 $\hat{Y}_w$ 是分亚区预报出来的,因此,利用(1)式可以估算出所预报地区各县(市)的单产,也可以求出各亚区的平均单产。至于(1)式中误差项 ε 的估计,一般可以根据未来的长期天气展望以及是否出现重大自然灾害等综合考虑。如无特殊情况,可以将 ε 取为0。

由于我们的样本有限,随着资料的延长,各主要参数如(3)式中的 Y_c, A, B , (10)式中的 b_{ij} ($j=0, 1, 2$)等都会有所变化。因此,原则上该预报模式只外推(使用)一年。每当有新的样本加入时,需修改相应的数据穿孔纸带,重新上机筛选因子,使所有参数取值得到调整。这样,可使该模式总是处于最佳状态。

参 考 文 献

- [1] 中央气象局农业气象研究室,编制棉花播种期预报的讨论,气象学报,31卷1期,1—9,1960。
- [2] 服永纲,编制冬小麦收获期预报的讨论,气象学报,31卷1期,36—40,1960。
- [3] 蓝鸿第,吕和同志农业气象预报经验的验证,大气科学,4期,312—315,1977。
- [4] 蓝鸿第等,东北地区粮豆产量预报,科学通报,27卷6期,383,1982。
- [5] 许尊伍,陕西汉中盆地水稻冷害初探,气象学报,40卷1期,89—95,1981。
- [6] 汤志成,介绍几种常用的农作物产量资料处理方法,农业气象科学,1期,92—96,1981。
- [7] 邓根云等,小麦分蘖的数学模式及其在栽培和农业气象中的应用,农业气象,总9期,15—19,1981。
- [8] 方开泰,聚类分析(I),数学的实践与认识,1期,66—80,1978。
- [9] 蓝鸿第等,长江中下游水稻日照温度农业气候指标的初步研究,气象学报,33卷4期,537—548,1963。
- [10] 中央气象局农业气象研究室等,冬小麦降水临界期的初步探讨,气象学报,31卷1期,51—59,1960。

- [11] 蓝鸿第等,我国水稻代表品种光照农业气象指标的研究,科学通报,23 卷 5 期,305—310,1978。
[12] 蓝鸿第,水稻发育速度模式的初步研究,科学通报,27 卷 3 期,180—185,1982。

A PRELIMINARY STUDY OF YIELD FORECAST MODEL FOR WINTER WHEAT IN NORTH CHINA

Wan Gongzhan

(*Meteorological Institute, Tianjin Weather Bureau*)

Wang Baocheng

(*Tanggu Meteorological observatory, Tianjin*)

Lan Hongdi

(*State Meteorological Administration, China*)

Abstract

The paper presents a crop-weather model for yield prediction on provincial level and higher. First, North China area is divided into a number of subregions by the R-type cluster analysis of climatic yield. Based on different effective of forecast period (EFP), correlations between climate yield and weather factor for each area are computed to sieve out forecast factors. Finally, forecast equations are established using the method of nonlinear multi-regression of accumulated and factorial type. The model may be applied to different crops, the EFP can be adopted according to characteristics of crops development stage or production needs.

The EFP of winter wheat yield forecast in Norther China is divided into three phases: before wintering, after wintering and before ripeness. The multiple correlation coefficients are larger than 0.70, 0.90 and 0.95 for the forecasts of the three different EFP respectively. The correlation error is below 15%, 8% and 4% respectively.