第 25 卷第 4 期　　南 京 气 象 学 院 学 报　　Vol. 25 No. 4
2002 年 8 月　　Journal of Nanjing Institute of Meteorology　　Aug. 2002

文章编号: 1000-2022(2002)04-0481-08

粮食生产风险水平的概率分布计算方法

邓　国,　王昂生,　周玉淑
(中国科学院 大气物理研究所, 北京　100029)

摘　要: 任意选取中国各地 40 a 以上粮食产量序列, 利用正态分布曲线、偏态分布正态化以及解析概率密度曲线积分等方法分别获得粮食产量序列的风险概率, 该结果反映了粮食生产不同增减产幅度及相应的概率大小。进行风险估计可以为发展生产、防灾抗灾提供参考。

关键词: 粮食产量; 概率分布; 风险水平
中图分类号: F307.11; X820.4　　文献标识码: A

农业灾害具有二重性: 自然属性和社会属性。农业灾害的自然属性, 主要是指致灾因子如洪涝、干旱、冰雹、冷害等危害农作物生长发育和产量形成, 破坏农业基础设施甚至造成人员伤亡等直接后果, 以及无家可归、停工停产造成经济损失等间接后果。当这些影响与公众期待的状况出现难以接受的差距时, 可能产生经济秩序混乱、民众恐慌等, 就上升为社会问题了[1]。因此, 粮食生产同时具有自然风险和社会风险。

风险是下面两个主要因素的函数: (1) 一事件或一系列不同强度事件发生的概率; (2) 事件的后果[2-3]。本文从粮食产量序列入手, 以概率统计方法为手段, 将风险分析理论引入农业气象灾害研究领域, 着重研究某一地区粮食产量序列风险水平的高低状况, 为科学、定量地反映该地区粮食产量的波动情况及其受气象灾害等不利因子影响大小提供依据。

1　粮食风险分析的基本理论

粮食风险分析是通过对农作物生长期内可能遭遇到的不确定事件的识别、衡量和处理, 以最小的成本将各种不利后果减少到最低程度的科学管理技术。粮食风险分析技术包括风险辨识、风险估计和风险评价与决策 3 部分[2-4]。其含义和关系可由图 1 表示。

1.1　风险辨识

风险辨识是对面临的潜在风险加以判断、归类和鉴定风险性质的过程。对粮食产量风险来说, 要求回答以下几个问题:

(1) 应该考虑哪些风险事件? 按照作物对不良条件反馈滞后时间的长短不同, 可将风险事件分为慢性发作型和急性发作型。前种风险事件, 是当不利条件对作物的抑制作用累积到相当

收稿日期: 2001-12-30; 改回日期: 2002-02-25
基金项目: 世界银行第 4 期对华技术合作 No. A3 "中国防灾减灾分析与对策"项目
第一作者简介: 邓　国(1971-), 男, 河北承德人, 博士, 从事农业气候资源区划和中小尺度灾害性天气数值模拟研究.

程度时才会对产量造成较大的影响; 后一种风险事件, 是使植株发生明显的病理变化或器官损伤, 如茎杆折断、叶片卷曲、落花落果、倒伏等造成的。

(2) 引起风险事件的原因是什么?致灾因子和孕灾环境是引发风险事件的原因[5]。例如植株叶片枯萎, 可能是由于大气干旱、土壤干旱、植株生理干旱等原因造成的, 而久旱不雨、地表蒸发过于旺盛、土壤中有毒物质的积累是更进一步的原因。探讨风险事件的原因可以在田间进行诊断, 赫德森提出一种产量变异推导法表[6](表 1) , 从作物发病(即风险事件) 的地理范围与产量年际波动等方面来推断作物产量变异的原因。

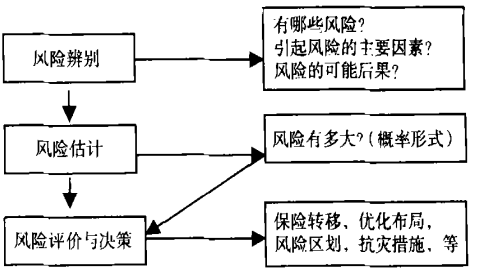


图 1 粮食产量风险分析内容

Fig. 1 The content of grain yield risk analysis

表 **1** 作物产量变异的原因

Table 1 The cause of grain yield variation

变异类型	原因
同一农场的田与田之间的差别(即同一管理水平之下)	由于土壤和小气候的差异
在类似土壤和同一地区农场之间的差别	由于经营管理能力的差异
同一田块不同年度之间的差异(同样的土壤和地理条件, 类似管理水平)	由于气候的差异

(3) 风险事件引起的后果及严重程度如何? 风险事件发生后使农作物减产是众多不利影响中最重要的部分, 同时还会对社会生产、生活造成更深远的作用。

1. 2 风险估计

风险估计是在风险辨识的基础上, 给出某一风险事件发生的概率及其后果[1]。但是, 农作物的全生育期可能遭受一系列致灾因子的作用, 而且这种作用是非线性形的, 难以分离出单一致灾因子的危害后果。概率分布是另一种风险估计的方法[2-4]。对给定的相对气象产量, 本文分别采用统计和积分的方法, 利用三种不同的方法可以计算出任意增产率区间和减产率区间的风险水平。

1. 3 风险评价与决策

风险评价与决策的内容是对于已知粮食风险判断是否采取措施, 采取什么措施, 以及采取措施后可能出现什么后果等作出决定的。风险评价的方法主要有如下几种[7-8]:

(1) 回避风险方法: 包括各种采用防灾、抗灾措施, 抗御不利气象条件对作物的危害。

(2) 权衡风险的方法: 针对局部区域风险水平的差异采取不同的农业技术措施, 包括相应的作物布局、耕作方法、管理方法、防灾、抗灾、救灾方法等, 使实际产量尽可能达到目标产量, 减少灾害的不利影响[9-10]。

(3) 风险评价的综合分析方法: 最优的农作物布局可以使粮食生产综合经济效益最佳, 考虑品种搭配、种植面积、播种日期、农业投入等因子可以实现这一目标。

(4) 保险转移: 由农民向保险公司投保, 以交纳保险费为代价将风险转嫁给保险公司, 一旦作物遭受的损失属于保险公司职责范围内, 则损失由保险公司予以补偿。

由图 1 可见, 在粮食产量风险分析中, 风险估计占据非常重要的地位。本文将风险理论引

入农业气象灾害研究中, 采用统计方法, 将粮食产量风险水平以概率的形式表示出来, 为政府部门决策提供依据。

2　资料处理

本文使用的粮食产量资料来自统计部门, 序列长度为 40 a 以上, 其中省级除天津和西藏外, 均为 1949—1996 年资料。

2.1　粮食单产分解

影响粮食产量形成的各种自然和非自然因素, 可以按影响的性质和时间尺度划分为农业技术措施、气象条件和随机 "噪声 "三大类。其中 "随机噪声 "类比较小, 可忽略不计, 粮食产量可分解为[11]

$$Y = Y_t + Y_w。 \tag{1}$$

其中 Y 为粮食单产, Y_t 是反映生产力发展水平的长周期产量分量, 即趋势产量; Y_w 是受以气象要素为主的短周期变化因子影响的产量分量, 即气象产量。若采用直线滑动平均模拟的方法求出趋势产量, 则气象产量为

$$Y_w = Y - Y_t。 \tag{2}$$

而

$$x = Y_w / Y_t。 \tag{3}$$

不受不同农业技术水平的影响, 称为相对气象产量。其物理意义表明粮食波动的幅值, 不受时间和空间影响, 具有可比性, 能较好地描述以气象要素为主的各种短期变动的因子对产量的影响。图 2、图 3 分别为河北省、河南省粮食相对气象产量序列图。从图中可以看出, 两地产量序列均在 60 年代初有较大的减产, 这与当时的社会历史背景和自然条件有关; 产量波动性在 80 年代以前比较剧烈, 而 80 年代以后波动幅度相对较小, 可能与中国旱涝灾害地域性分布的变化有关[12]。

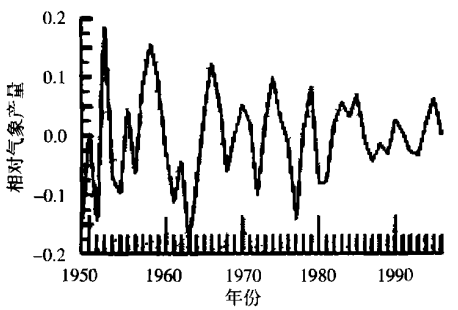


图 2　河北省粮食相对气象产量序列

Fig. 2　The series of relative meteorology yield of Hebei Province

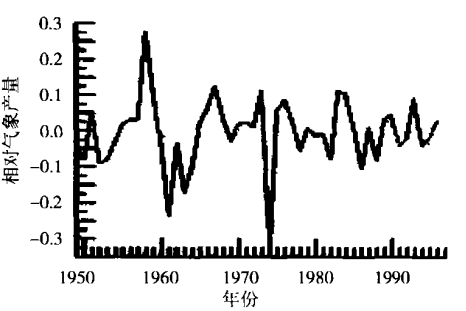


图 3　河南省粮食相对气象产量序列

Fig. 3　The series of relative meteorology yield of He´nan Province

2.2　正态性检验

考虑到作物产量形成过程中其不确定性主要由各种气象因子的波动造成的, 而有些气象因子如气温等时间序列符合正态分布, 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; 与此类似, 农作物产量也表现为平年较多, 大丰大欠年份较少。由此推断, 粮食序列产量的波动可能也具有正态性的分布特点。

考虑到影响粮食产量变化的某些气象因子具有正态性的特点, 为验证消除了生产力水平影响的相对气象产量是否也具有正态性, 采用偏度—峰度联合检验法和 χ^2 检验法[13-14] 分别进行了检验, 经计算, 河南省粮食单产相对气象产量序列没有通过正态性检验; 河北省粮食单产相对气象产量时间序列通过了检验, 即可认为河北省相对气象产量序列符合正态分布。

3 粮食产量序列的风险估计

风险分析最重要的目的是回答实际产量与目标产量偏离不同幅度出现的概率。本文的目标产量即趋势产量, 偏离趋势产量的正负值的百分比即增减产百分比。

3.1 正态型相对气象产量序列的风险估计

对已经通过正态性检验的相对气象产量序列, 则可认为该序列的概率密度曲线具有正态分布的特点。根据正态分布曲线的性质, 很容易获得粮食产量的风险水平。

若已知正态分布函数的数学期望为 U, 均方差为 Σ, 则该正态分布函数的性状就已确定。但是, 事实上是不可能得到相对气象产量序列总体的这两个特征参数值的。实际工作中常用大样本平均值 u 代替总体数学期望 U, 样本均方差 σ 代替总体均方差 Σ(在农业气象研究中, 所谓大样本是指样本数 N 大于 30)[15]。本文粮食单产资料基本上可以满足这种近似要求。因此, 可以用样本平均值 u 和样本均方差 σ 来建立概率分布密度函数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2\sigma^2}(x-u)^2]。 \tag{4}$$

分布函数为

$$F(x) = \int_{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2\sigma^2}(z-u)^2]\mathrm{d}z。 \tag{5}$$

损失概率和损失严重程度是风险估算所要解决的问题。因为相对气象产量序列表示粮食产量波动的情况, 所以可以利用分布函数来估计水稻产量不同增产或减产幅度对应的概率大小, 从而给出未来年份某一地区水稻的风险状况[2-4]

$$\int_1^2 f(x)\mathrm{d}x = p(x_1 < x \leq x_2)。 \tag{6}$$

损益幅度小于 x_0 的概率

$$p(x < x_0) = \Phi(\frac{x_0-u}{\sigma})。 \tag{7}$$

损益幅度大于 x_0 的概率

$$p(x > x_0) = 1-\Phi(\frac{x_0-u}{\sigma})。 \tag{8}$$

损益于 x_1 和 x_2 之间的概率

$$p(x_1 < x_0 \leq x_2) = \Phi(\frac{x_2-u}{\sigma}) - \Phi(\frac{x_1-u}{\sigma})。 \tag{9}$$

上面所说的 “损失风险 ”, 不仅包括粮食减产的情况, 对增产也有同样的含义。比如如果未来年份粮食生产获得大丰收(增产幅度> 30 %) 的风险概率为 10 %, 而粮食产量低于这个水平的风险概率高达90 % 。河北省相对气象产量序列符合正态概率密度曲线, 经计算粮食减产风险概率大于 10 %的概率为 11.3 % (即相对气象产量 x 小于- 10 %)。

3.2 偏态分布正态化的方法估计风险水平

对未通过正态性检验的相对气象产量序列, 本文采用一个简化的方法, 即 “偏态分布正态

化 ”, 通过将普通坐标转化为几率坐标, 使在普通坐标上呈非正态分布的数据, 转换为正态分布。这种方法普适于分析样本序列本身波动性较大, 或者序列中资料的分布范围太广, 或者资料的均匀性并不十分理想的情况[16]。正态化之后, 就可以得到相对气象产量序列的特征参数平均值与标准差, 为以后的分析打好基础。

偏态分布正态化的基本步骤如下:

(1) 将相对气象产量序列$\{x_i\}$分为若干组, 分组数为 $m = 1.52(n-1)^{2/5}$, n 为样本数。本文 n 为 48, 则分组数为 7。

(2) 将样本数据由小到大顺序排列, 找出最小值 x_{min} 与最大值 x_{max}, 求出极差 R。根据极差与分组数据测定组距、组界值。组距 $\Delta x = R/m$, 为了将 x_{min} 与 x_{max} 包括在分组内, 横坐标取值下限应小于 x_{min}, 取值上限应大于 x_{max}。在确定组界时, 为了避免位于界点上的测定值在分组时发生跨越组的问题, 组界值的有效数字应比原始测定值多取一位, 该位的值取 0.5。

(3) 计算各组的频数和相对累积频数, 由标准正态分布表求得概率坐标。由相邻两组概率坐标相减求出概率组距。

(4) 计算坐标转化后的概率频数 $f =$ (组频数/概率组距), 总频数 $F = \sum f$。

(5) 计算相邻概率坐标的平均值。

(6) 计算正态化后的平均值 u 与标准差 σ。

经 “正态化 ”处理的相对产量序列就可以视为 “正态序列 ”, 以新的均值和均方差作为特征参数进行风险估算。

3.3　一般相对气象产量序列风险概率的计算方法(解析概率密度曲线法)

对任意一个相对气象产量序列, 为估计其风险水平, 需分以下几个步骤进行。

3.3.1　用近似解析式表示相对气象产量序列的概率密度曲线

在相当长的一段时间内, 由于每年的气象条件、农业生产措施等不同造成粮食产量的不确定性, 使相对气象产量序列具有随机变量序列的特点(图 4), 其数值一般介于 -1 和 1 之间。为研究该随机序列的特性, 本文采用近似解析式来构造相对气象产量序列的概率密度函数, 将看似无序的相对气象产量序列的性状表示出来。

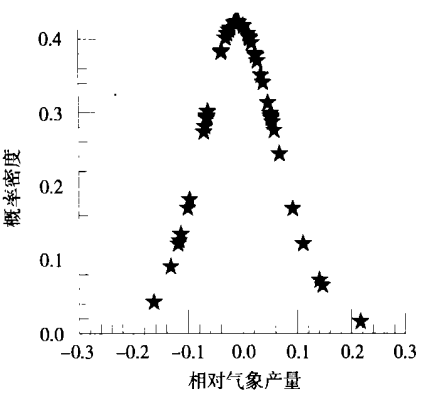


图 4　河北省相对气象产量正态型概率密度曲线

Fig. 4　Normal probability density curve of relative meteorological yield in Hebei Province

令 $f(x)$ 是相对气象产量序列的概率密度函数的解析式形式[17]

$$f(x) = \frac{1}{2\pi}\exp\left[-\frac{x^2}{2}\right]\left[1 + \sum_{n=3}^{\infty}\frac{c_n H_n(x)}{n!}\right]。\tag{10}$$

其中 $H_n(x)$ 为艾尔米特多项式, 可表示为

$$H_n(x) = x^n + \sum_{m=1}^{n/2}\left[(-1)^m(2m-1)!!c_n^{2m}x^{n-2m}\right]。\tag{11}$$

系数 c_n 与概率密度 $f(x)$ 所表征的相对气象产量序列 x 的 k 阶矩有关。本文以河北省粮食产量序列为例, 系数取前 6 项, 分别为 $c_1 = 0$, $c_2 = 0$, $c_3 = 0.109\ 62$, $c_4 = 0.354\ 102$, $c_5 = 0.323\ 049\ 7$, $c_6 = -0.658\ 931$; x 为正规化的相对气象产量序列。利用(10)式只需要不多几项就可精确地表

示相对气象产量这种随机变量的的概率密度函数。

根据(10) 式, 本文以省级粮食产量序列为例, 得到河南省粮食产量序列概率密度曲线, 该曲线描述了相对气象产量序列的统计性质, 如图 5 所示。从该曲线可以看到, 相对气象产量的增减产概率主要集中于– 0. 3～0. 3 之间, 还可见到, 概率密度曲线为偏态的, 与上文偏度—峰度检验的结果相吻合。这表明采用解析式方法刻画相对气象产量序列的正确性。概率密度曲线的解析式虽然从离散点拟合而来, 但该解析式同样适用于一般相对气象产量序列的变化范围, 即– 1～1 之间。即通过该式, 可以将研究离散的相对气象产量序列的性质转化为对连续性随机变量 $f(x)$ 的研究, 不但可以得到已有数据的统计性质, 也可以推断一般增产率区间或减产率区间的性质。

3. 3. 2 高斯求积法计算分布函数

从相对气象产量序列得到的概率密度曲线 $f(x)$ 可以视为连续型随机变量的概率密度函数, 为研究在不同增产率区间和减产率区间的风险概率, 需要从概率密度曲线构造分布函数。在相对气象产量变化区间的分布函数为

$$F(x) = \int_{-\infty} f(t)\,\mathrm{d}t。 \qquad (12)$$

即分布函数的计算归结为积分的计算。

由于高斯求积法是具有最高代数精度的求积公式[18], 故本文采用该法计算分布函数。对定积分

$$F(x) = \int f(x)\,\mathrm{d}x \approx \sum_{k=1}^{n} [B_k f(x_k)], \qquad (13)$$

如果结点 x_k 不固定, 把它看成 n 个可自由选择的参数, 则上式右端有 $2n$ 个 B_k 和 x_k 待定参数, 那么该公式就可能对 $2n-1$ 次多项式精确。如何确定系数 B_k 及结点 x_k, 使上式代数精度达到 $2n-1$, 这就是构造高斯高精度求积公式的基本思路。

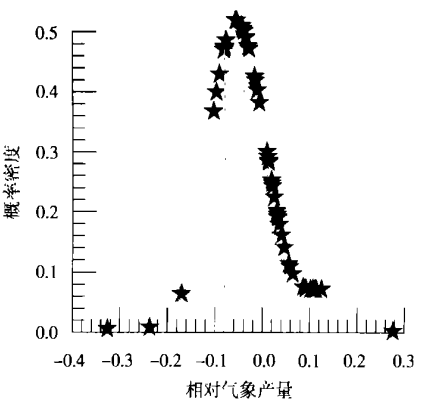


图 5 河南省粮食产量序列非正态概率密度曲线

Fig. 5 Non-normal probability density curve of grain yield series in He´nan Province

分布函数为非降函数, 在任意一点上, 表示相对气象产量小于该点的情形, 但为了直观表示相对气象产量实际增减产幅度, 本文所求的概率密度曲线以相对气象产量的变化范围作为 x 轴, 故概率密度曲线所围面积不等于 1, 即分布函数 $F(x) = \int_{-\infty}^{\infty} f(t)\,\mathrm{d}t \neq 1$, 尚需做相对化处理。

3. 3. 3 计算任意增产率区间或减产率区间的风险概率

由图 4、图 5 可见, 对任意相对气象产量序列, 表示所有可能生产状况可以用分布函数 $F(x) = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x$ 来表示。对某种生产状况, 如粮食生产增产率或减产率介于 a 与 b 之间可以用分布函数 $\int f(x)\,\mathrm{d}x$ 来表示。则介于 a 与 b 之间的风险概率

$$p(a < x < b) = [\int f(x)\,\mathrm{d}x] / [\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x]。 \qquad (14)$$

4　三种粮食产量序列风险水平估计方法结果比较

本文利用河北、河南两省粮食单产资料, 分别计算在不同增产率和减产率区间的风险概率。根据上面结果, 知道河北省相对气象产量序列符合正态分布, 而河南省则是非正态的。下面采用不同的方法计算两省不同增产率区间和减产率区间的风险概率(表 2) 。

表 **2**　不同风险估计方法计算结果比较

Table 2　Comparision of results calculated by different risk evaluations

增产率区间或减产率区间	河北		河南	
	解析概率密度曲线	正态概率密度曲线	解析概率密度曲线	正态概率密度曲线
- 30 %～- 20 %	0. 7	0. 6	1. 2	2. 8
- 20 %～- 10 %	10. 6	11. 3	18. 3	20. 1
- 10 %～- 5 %	16. 8	16. 1	27. 6	20. 7
- 5 %～0 %	23. 9	23. 2	26. 0	22. 6
0 %～5 %	23. 2	23. 5	13. 0	17. 8
5 %～10 %	15. 3	15. 7	5. 1	10. 1
10 %～20 %	9. 0	8. 9	6. 9	5. 4
20 %～30 %	0. 5	0. 5	1. 2	0. 3

由表 2 可见, 河南省粮食产量风险水平高于河北省, 表现为在增产或减产幅度高的区域风险概率大。在各个风险水平上, 符合正态分布的相对气象产量序列的风险水平和按照积分方法得到的结果非常接近。这说明, 两种方法对于计算粮食产量风险水平都很有效, 两者相互印证。而采用偏态分布正态化所得结果与积分方法有一定出入, 说明偏态分布正态化的方法相对较粗糙。一般相对气象产量序列风险水平的求算方法不需要进行检验, 但该序列必需具有表征数字特征的各阶矩; 对于符合正态分布的相对气象产量序列采用正态曲线求风险水平的方法简单明了; 采用偏态分布正态化的方法精度最低, 主要是该方法的实质在于通过坐标转换使在一般坐标下不符合正态分布的数据转化为其他坐标符合正态分布的数据。

5　结论与讨论

(1) 在计算粮食生产风险水平时, 以风险分析为指导思想, 采用数理统计方法, 试图提出三种根据粮食产量序列计算粮食生产风险水平的方法: 正态型相对气象产量序列的风险水平估计方法、非正态型粮食产量序列采用偏态分布正态化的方法和解析概率密度曲线的风险水平估计方法, 这些方法在粮食产量风险水平评估中目前尚未见到。

(2) 风险分析最重要的目的之一就是要回答实际产量偏离目标产量一定正负幅度出现的概率, 本文的目标产量即趋势产量, 偏离目标产量的正负值即增减产百分比。由直线滑动平均法可求出目标产量; 在样本数目充分大的情况下, 样本的特征接近总体的特征, 借助于概率统计的方法把样本离散点的描述变成对样本总体的连续性变量的描述, 从而对实际产量资料序列的性状描述上升到了理论的高度。

(3) 本文用定量化的方法估算了任意粮食产量序列的风险水平, 结果较好地反映了粮食生产波动的状况。随粮食产量资料样本的增加, 结果更加稳定可信。三种方法计算的结果相互具有可比性, 其模型具有普适性; 结果可为决策者防灾救灾、发展生产提供参考。

参考文献:

[1] William J Petak, Arthur A Atkisson. 自然灾害风险评价与减灾政策[M]//向云立, 等译. 北京: 地震出版社, 1993: 7–16; 63–68.
[2] Stahl Michael J. Strategic management for decision making[M]. Boston: PWS–Kent Pub. Co., 1992: 167–172.
[3] 风险管理编写组. 风险管理[M]. 成都: 西南财大出版社, 1994.
[4] 郭仲伟. 风险分析与决策[M]. 北京: 机械工业出版社, 1987.
[5] 郑有飞, 万长建, 缪启龙, 等. 气候变化对小麦生产影响的数值模型研究[J]. 南京气象学院学报, 1999, 22(4): 645–651.
[6] 刘乃壮. 作物气象病害与防治[M]. 江苏: 江苏科学技术出版社, 1992: 28–30.
[7] James L Athearn. Risk and Insurance[M]. 4th Ed. St Paul: West Publishing Company, 1984: 7–12.
[8] 潘履夫. 保险学概论[M]. 北京: 中国经济出版社, 1995.
[9] 冯定原, 邱新法, 杨幸福. 农业气候资源影子价格初探[J]. 南京气象学院学报, 1994, 17(4): 432–437.
[10] 李湘阁, 何海燕, 景元书, 等. 水稻旱育抛秧生长发育的动态模拟模型[J]. 南京气象学院学报, 1999, 22(4): 587–595.
[11] 王馥棠, 李郁竹, 王石立. 农业产量气象模拟与模型引论[M]. 北京: 科学出版社, 1990: 40–41; 53–61.
[12] 彭加毅, 孙照渤, 朱伟军. 70 年代末大气环流及中国旱涝分布的突变[J]. 南京气象学院学报, 1999, 22(3): 300–304.
[13] 方开泰, 许建伦. 统计分布[M]. 北京: 科学出版社, 1987: 136–152; 331–332.
[14] 全国统计方法应用标准化技术委员会标准化中统计方法应用分委员会. 数据的统计处理和解释—— 国家标准中若干统计方法[M]. 杭州: 浙江大学出版社, 1991.
[15] 魏淑秋. 农业气象统计[M]. 福州: 福建科学技术出版社, 1985: 12–13.
[16] 邓 勃. 分析测试数据的统计处理方法[M]. 北京: 清华大学出版社, 1995.
[17] Åra eB B C. 随机函数论及其在自动控制中的应用[M]//田欣为, 等译. 北京: 科学出版社, 1966: 17–47.
[18] 葛孝贞. 大气科学中的数值方法[M]. 南京: 南京大学出版社, 1994: 204–222.

Grain Yield Risk Level Calculated by Probability Distribution

DENG Guo, WANG Ang–sheng, ZHOU Yu–shu
(Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China)

Abstract: There have been lots of methods on the prediction of grain yield, however these predicted results could not be very accurate due to all kinds of factors that affect on grain yield. In this article, three different kinds of theoretic probability methods (i.e, the normal distribution curve probability, converting the abnormal distribution curve to normal curve, the integration of analytical probability density curve are used to calculate the probability of risk of any given grain yield sequence. These methods showed a wide range of increase and decrease of the crop yield as well as the corresponding high and low probability, which can provide good references to the development of production, natural disaster prevention and alleviation.

Key words: grain yield; probability distribution; risk level