●于系民　赵焕宸

蒙特卡罗方法在农业气象学中的应用

气象出版社

蒙特卡罗方法在农业气象学中的应用

于系民　赵焕宸　　著

气象出版社

内 容 提 要

本书论述了蒙特卡罗方法用于农业气象学的意义、途径和方法步骤。内容包括对农业气象回归模式的蒙特卡罗检验、灾害风险分析、小气候改造最优方案的选定、天气-畜病扩散预报模型、农业气象随机服务系统与库存系统。

本书主要供农业气象研究人员阅读。对农林牧气象方面的技术工作者和有关专业的大专院校师生，也有较好的参考价值。

蒙特卡罗方法在农业气象学中的应用

于系民　赵焕宸　著

责任编辑　黄　健

• • •

气象出版社出版

（北京西郊白石桥路46号）

北京怀柔燕文印刷厂印刷

气象出版社发行　全国各地新华书店经售

• • •

开本：787×1092　1/32　印张：2.75　字数：56千字

1989年11月第一版　1989年11月第一次印刷

印数：1—1000　定价：2.00元

ISBN 7-5029-0271-6/S·0030

序

在理论农业气象学的研究中，时而碰到用经典数学模型无法求解的难题。一次偶然的机会，蒙特卡罗方法被我们引进农业气象学，没想到，这一初步尝试竟迅速地得到同行的注意，几篇论文先后被国际学术刊物和特种学术文献采用，一些学术单位请作报告或举行专题讲座，有些气象台站则很快用于业务实践。

听讲者和实用者希望早日拿到文字材料，从事研究的专家则敦促尽快出书以飨读者，此即该长篇论文以本书形式问世的基础背景。

本书的内容主要是笔者近年的研究成果。该书在阐明农业气象学随机模拟、系统仿真和蒙特卡罗方法应用于农业气象学的基本原理和方法之后，讨论了其渗入农业气象学各个分支（含农业气象运筹优化、农业气象经济学等）的实例。

因论题较新，故尚无可借鉴的同类书籍，书中缺点错误在所难免。依“宁献丑，勿藏拙”的治学策略，将论文发表出来，作为引玉之砖。倘有读者，在读本书前，对“蒙特卡罗方法在农业气象学中的应用”这一论题不甚了解，读后却乐于介入其研究或实用之行列，促进它不断发展，那就可以说笔者的目的已达到，将以此深感欣慰！

最后我们诚恳地感谢：康拉德（维也纳大学）、史密斯（英国气象局）、雷芙斯奈德（耶鲁大学）、蓝鸿第（气科院）、张建中（计算中心）和郭景（沈阳计算所）诸先生，他们分别同笔者作过有益讨论，审阅过有关文稿并推荐出版。

于永民、赵焕宸

1987年10月1日

引 论

许多学者善于用"一种科学只有在成功地运用数学时，才能达到真正完善的地步"（马克思语）这句名言指导数学在其他学科中的应用，从而丰富各学科的理论体系和方法。各学科大都在探索定量揭示事物本质的途径。既属于农业基础科学又属于应用气象学的农业气象学，在其漫长的发展史中，也在逐步深入地运用数学完善其自身。近年，特殊函数、模式识别以及超越微分方程的数值解法在农业气象学中的应用，实现了农业气象场（而非由点推面）的预报，促进了农业气候分析评价的定量化、模式化和自动化。

数学本身也在不断向深广发展，农业气象学正持续地从中汲取养分。农业气象学所用到的数学，不应局限于数学分析、数学物理方程、概率论与数理统计、运筹学、模糊数学与控制论，而应考虑其他数学分支。农业气象学中开始应用的蒙特卡罗方法（Monte Carlo Method），其原始思想产生得很早，用频率近似地估计概率的随机试验方法，实际上是古代的蒙特卡罗方法。但用这种方法真正解决问题，是现代的事。近几十年来，随着电子计算机的出现和发展，人们才有意识地、广泛地、系统地应用它解决物理学、生物医学等方面的大量问题，并把它作为计算数学的一个重要分支。这是一种具有独特风格的数值计算方法，它既能求解确定性的数学问题，也能求解随机性的数学问题。而农业气象数学模型恰好涉及这两种类型的数学问题。于是我们考虑"蒙特卡罗方法在农业气象学中的应用"这一论题。

作为本研究的基础，让我们首先讨论农业气象综合因子（agrometeorological synthesis）所致生物增长的确定性数学模型问题，其次阐述与它有关的若干随机模拟问题。这里所谓农业气象综合因子系指生物个体或群体某一发育阶段或整个生命史期间各种农业气象条件（温度、湿度与光照等）的有机综合。这里所涉及的农业是包括种植、林业、牧业在内的大农业，故生物指动植物和微生物。鉴于各种生物其生命史上的差异和研究目标的区别，农业气象综合因子所涉及的时间范围是变化的。例如，属于禾本科植物的大田作物是一年生的，可求其全生育期或个别发育期（如苗期）的农业气象综合因子；对于多年生的果树可求盛果期某年的农业气象综合因子；对于家畜，可求幼龄期农业气象综合因子；对于食用菌，可求真菌群体农业气象综合因子等等。农业气象综合因子的求法，可用试验法、解析法、概率统计法等等，此不赘述。

令 $x(t)$ 为实值的连续函数，它表示 t 时刻农业气象综合因子所致生物群体量（这种量可以是植物的、动物的，也可以是微生物的。为简便起见，这里用微生物的，以黑木耳真菌群体中的农业气象综合因子所致的真菌数目为例）。为了描述真菌群体量的增长动态，必须根据某些假定条件来构造数学模型。我们在这里假定：（1）在时刻 t，群体中有 x 个真菌（均指农业气象综合因子所致，下同）；（2）群体中真菌的含量只能增加，并且在有限的时间区间 $(t, t+\Delta t)$ 内的增长量与时刻 t 的真菌数目 $x(t)$ 成正比。于是，我们有

$$\Delta x(t)=\lambda x(t)\Delta t, \qquad \lambda>0,$$

由此导出微分方程

$$\frac{dx(t)}{dt}=\lambda x(t) \tag{1}$$

若假定$x(0)=x_0>0$，则式（1）的解为

$$x(t)=x_0\exp(\lambda t) \tag{2}$$

在上述的简单模型中，未假定群体中真菌数目的减少。因此，由式（2）可知：t增大时，群体的真菌含量增加，$t=0$时真菌含量为最少，即$x(0)=x_0$；只要初值x_0相同，则在给定时刻$t>0$，真菌含量总是相同的。

然而，农业气象综合因子一般是随机的，真菌含量也并不是永远增加的。所以应考虑与上述模型相应的随机模拟问题，从而建造随机模型。设整数值的随机变量$X(t)$表示时刻t群体中的真菌数目，并假定$X(0)=x_0>0$。我们要研究的问题是，在时刻t，真菌数目等于x的概率$P_x(t)=\mathscr{P}\{X(t)=x\}$。

为了建造农业气象随机模型，根据生物群体的随机增长规律和生物数学基本规律，我们假定：

（1）如果在时刻t，群体中有$x>0$个真菌，那末在时段$(t, t+\Delta t)$内，由农业气象条件而增加一个真菌的概率为

$$\lambda x\Delta t+0(\Delta t),\ \lambda>0;$$

（2）在$(t, t+\Delta t)$内，群体中增加两个或两个以上真菌的概率为$0(t)$。

根据A.T.Bharucha-Reid在《马尔可夫过程论初步及其应用》一书中的推导结果，有关系式

$P_x(t+\Delta t)=(1-\lambda x\Delta t)P_x(t)+\lambda(x-1)\Delta tP_{x-1}(t)+0(\Delta t)$ 当$\Delta t\longrightarrow 0$时，得方程

$$\frac{dP_x(t)}{dt}=-\lambda x P_x(t)+\lambda(x-1)P_{x-1}(t),\quad x=x_0,\ x_0+1,\ \cdots\cdots \tag{3}$$

该方程求解的初条件为

$$P_x(0)=1,\quad x=x_0\text{时},$$
$$=0,\quad \text{在其他条件下} \tag{4}$$

对$x\geqslant x_0$，方程（3）的解为

$$P_x(t)=\mathscr{P}\left\{X(t)=x\right\}=\binom{x-1}{x-x_0}\exp(-\lambda x_0 t)(1-\exp(-\lambda t))^{x-x_0} \tag{5}$$

这种概率表达式即是随机模拟的结果。蒙特卡罗方法是一种随机模拟方法。它在农业气象学中应用的基本思想是：为了求解农业气象学原理、农业气象预报与情报、农业气候以及农业气象随机服务系统等方面的问题，首先建立一个随机过程，使其参数等于问题的解；然后通过对过程的观察或抽样试验来计算所求参数的统计特征，最后给出近似解。解的精度则用估计值的标准差表示。

虽然蒙特卡罗方法可求解确定性和随机性两种类型的问题，但在农业气象中所遇到的大都是随机性的问题，如气象灾害对农作物危害程度问题，农业气候条件对各种作物收获量的保障率问题，农业气象随机服务系统中的排队论问题，库存模型问题等等。因而，本文以处理随机性问题的蒙特卡罗方法在农业气象学中的应用为主。

在用蒙特卡罗方法解决农业气象问题的过程中，主要侧重以下内容：对求解的问题建造简单而又便于实现的概率统计模型，使所求的解恰好是所建模型的概率分或数学布期

望；根据概率统计模型的特点和实际计算的需要，尽量改进模型，以减小方差、降低费用；建立随机变量的抽样方法，包括建立产生伪随机数的方法和建立对所遇到的分布产生随机变量的随机抽样方法；给出获得所求解的统计估计值。

蒙特卡罗方法用于农业气象学本身的两个特有的不能为其他模型所代替的优点是：它用数值试验代替田间或人工气候室难于完成的真实试验；它用模拟方法求客观概率以代替以往的主观概率，因而可谓农业气象从主观向客观迈进的措施之一。

目　录

第一章　农业气象经验回归模型显著性检验的蒙特卡罗模拟方法

1. 引言

在农业气象学中，经验回归模型是得到广泛应用的一种有力的数理统计工具，无论是在理论研究工作中，或者是在实际业务工作中，该模型都被广泛地应用于作物气象学、畜牧气象学、农业气象预报情报以及农业气候等学科之中。象筛选多重线性回归这样的模型，在农业气象学中的应用尤其频繁。有了与模型相应的软件后，用起来是十分方便的。

近年来，有关基础学科与应用学科的研究者认真地讨论了模型检验问题。事实表明：在农业气象的回归分析计算中，需要进行预报与控制效果可靠性、稳定性的检验。一般的做法是，将给定的观测数据分成两个部分：一部分（占有大量的数据）用来建立回归方程；另一部分（占有少量的数据）用来检验已经建立的回归方程。数据量不大（不超过30）时，用这种方法就有困难了。在农业气象试验中，或在其他农业业务中，搜集到的农业气象数据，不足30的较多。这是因为农业气象需要的是平行观测，要有生物与气象两方面的系统记录，前一种记录一般比后一种少，如物候观测我国是按规范观测的少有的国家，但超过30年的观测资料至今尚少；此外，农气研究中，要求气象以外的其他条件的变化小，而又要同时反映当时生产水平，才能使农业实践用得上，这种矛盾很难解决，因为农业生产条件连续几十年保持稳定是不可能的。基于上述理由，在农业气象学中，急需一种能较客观地检验回归模型显著性的方法，我们在本章即通过一个实

例简要地阐明这种方法。

2.作为实例的问题

我们有自1958年以来的25年的谷子受粘虫为害的资料，而据理论与实测研究，粘虫迁飞同高空气流关系密切。于是想建立一个粘虫害同天气形势关系的经验回归模型

$$\widehat{Y}(t)=\sum_{i=1}^{m} B_i X_i(t)+A \qquad (1.1)$$

Y（t）代表粘虫害，A为回归方程的常数项，X_i（t）是自变量，m是方程的元数。这里X_i（t）是表示500hPa高空流场形势的量，它与Y（t）都是时间的函数。

利用历史上有关年份的高空形势图，我们可以取得有一定生物物理意义的作为时间函数的20多个备选自变量，利用这些量，我们用逐步回归法很容易作出方程，其形式如式（1.1）。我们要研究的是：方程(1.1)在统计上是否显著，对此加以检验。

我们在这里简单讲述一个必要的概念——因变量方差减小的百分率PR，其定义式是

$$PR=\frac{1-\sum_{1}(Y_i-\widehat{Y}_i)^2}{\sum_{1}(Y_i-\overline{Y})^2}\times 100 \qquad (1.2)$$

这里Y是观察到的粘虫害，$\widehat{Y}$和$\overline{Y}$分别是用模型估计的与求平均的相应数据。由试验得知，在5个自变量后，PR较稳定，所以作检验时取式（1.1）中的m＝5最为合适。5个自变量时，PR＝68，以后每增加一个自变量，PR至多增加3。

3. 经典检验法不适用于本问题的理由

在数理统计学中，经典的多元分析原理指出：对于一组呈正态分布的备选自变量来说，从中挑选自变量的过程中，用 F 检验法来检验其显著性，是很合适的，当各个备选自变量之间是不相关(确切地说是关系很不密切)的时候，从许多备选自变量中进行筛选以后，其结果一般是可用的。然而，当我们所假定的自变量与因变量之间实际上没有关系，而方程中的自变量却是从相互之间有关的备选自变量中筛选出来的时候，F的分布函数则是未知的。事实上，在我们所讨论的粘虫害与大气环流中有关物理量的关系这一问题以及好多同类的农业气象学问题中，尽管用于表达大气环流特征的物理量很多，如欧亚大陆环流分型(东方型、西方型、纬向型)、500 hPa 大气环流指数和副热带高压位置等等，它们之间的关系，就是单从物理意义上来看，也是很密切的；而与此同时，对于粘虫害与大气环流表征物理量之间的关系，直到目前还没有很充分的有明确的生物物理意义的解释，长期预报尤其如此。这种关系只能建立在统计基础上，备选自变量与因变量的实际关系，可能是不密切的．上述事实，已经说明，在我们提出的问题中，利用经典的F检验的不可靠性

此外，在我们所要研究的问题中，可供使用的粘虫害数据只有25年，假定它们是抽取自同一母体的样本，实际上仍是一个小样本。这就决定了数据不能分为“用于建立回归方程”的和“用于检验回归方程”的两部分数据。只有待到抽自同一母体的样本数据相当充足时，或依据方程建好之后新增加的信息来检验已作方程．才是适用的好方法。

目前，我们面临的重要问题是：依粘虫害与天气形势图

提供的数据，用逐步回归法作出的方程式，在统计上是显著的吗？不用F检验法，如何检验呢？下面讨论这些问题。

4.蒙特卡罗模拟检验的方法与步骤

（1）虚假设与方程舍用的标准：

本研究试用如下虚假设回答上面提出的问题，即：

用实际数据求出的粘虫害方差减小的百分率，与用随机选择的伪粘虫害所对应的方差减小百分率相等。

如果求出的数值差异显著，则拒绝上述虚假设，我们的结论是：接受检验的那个方程，是说明粘虫害与高空气流形势之间客观关系的方程；如果虚假设未被拒绝，说明方程不可用，必须另作方程。

使因变量方差减至68％的方程（1.1），被选用作检验。我们将此方程同依据随机选择的粘虫害数据而提出的伪方程，加以比较。在引出伪方程时，备选自变量是不以任何方式改变的。

（2）伪方程制作：

①正态随机数：为制作粘虫害依赖于500hPa形势的伪回归方程，须用正态分布的随机数作为伪因变量。但是，"如果不加特别说明，凡是说"随机数"，就是指[0，1]区间上均匀分布的随机数。"那么，怎样由一般随机数表的随机数转换出正态分布N（0，1）的随机数呢？王梓坤和Голенко专著指出：设η_1，η_2，…，η_n为n个独立的随机变数，且它们是均匀的，由于$E\eta_i=1/2$，$D\eta_i=1/12$，根据中心极限定理，随着k值的逐渐增大，$\xi k=\left(\sum_{i=1}^{k}\eta_i-\frac{k}{2}\right)\Big/\sqrt{\frac{k}{12}}$

的分布渐近于标准正态分布N（0，1），可把ηk近似取作N（0，1）的随机数。根据王梓坤《概率论原理及其应用》（253—266页)所述k取值12最为合适，如果计算设备较差，用上述近似方法即可。但为更精确，我们又取两独立的均匀变数η_1和η_2，作变换

$$\xi_1=(-2\ln\eta_1)^{\frac{1}{2}}\cos2\pi\eta_2 \qquad (1.3)$$

$$\xi_2=(-2\ln\eta_1)^{\frac{1}{2}}\sin2\pi\eta_2 \qquad (1.4)$$

则所得的ξ_1和ξ_2为两个相互独立的正态分布随机变数。著者在用式（1.3）和（1.4）时发现，一般随机数表中η_1和η_2均为两位以上整数，代入式（1.3）和（1.4）前可先变为小数，如389变为0.389，这样根号内数恒为正值。

②随机数-虫害级别图的制作：将25年的粘虫害实测值，按由小到大的次序排列起来，并用Kimball在“概率格纸点绘位置选择”(见《美国统计学会会刊》55卷546—560页，1960年）一文中的公式

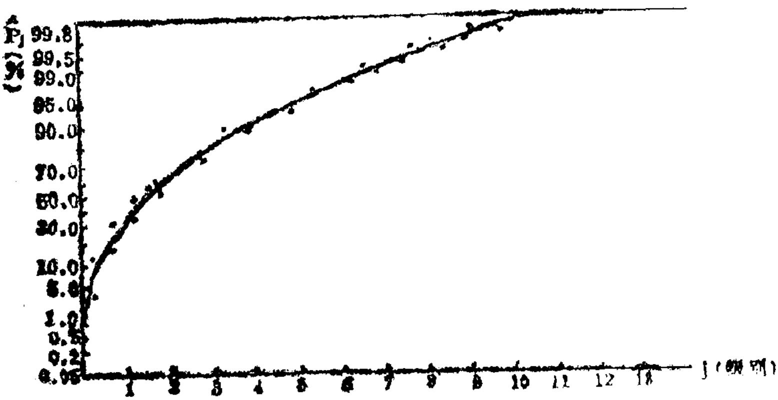

图1.1　用正态分布随机数查算伪粘虫害图解

$$\widehat{P}_j = \left(j - \frac{3}{8}\right) / \left(n + \frac{1}{4}\right) \qquad (1.5)$$

算出不同的j（观测到的虫害级别）对应的概率$\widehat{P}_j$，然后在通用的概率格纸上点绘成如图1.1所示的概率曲线。

③伪方程作法：首先利用①所述方法产生20组正态随机数，每组含25个；其次，用②的方法，查图1.1，将所有的随机数都转换成伪粘虫级别；用每组随机数对应的25个伪粘虫害数字，被用以作为Y（t）的观测值，以原有的500hPa天气形势表征备选自变量仍作为X（t）观测值，应用逐步回归筛选的程序作出五元一次线性方程，共作出这样的伪方程20个。

（3）伪方程的检验：

作出20个伪粘虫害依500hPa天气形势的方程后，先用式

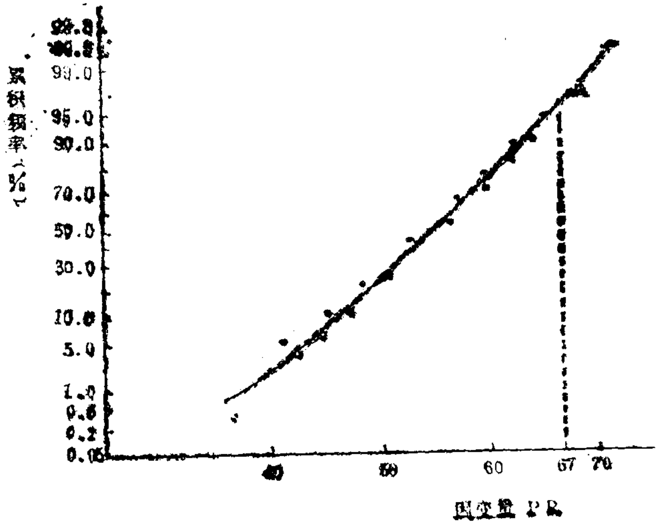

图1.2 用伪粘虫害气象综合因子求得的方差减小百分率PR之频率分布

图中△为用实测记录得出的结果

（1.2）求出各方程因变量方差减小的百分率PR，共20个，结果为40，41，43，58，56，…，65，71。

再将上述20个数字绘成如图1.2所示的累计频率图，按频率95%的横线为准，当用实际数据配合的方程对应的PR大于67时，应拒绝虚假设。本例中，用实际数据作出的方程PR＝68＞67，所以拒绝虚假设，我们接受粘虫害与高空气流关系的方程，具体经验式是：

$$Y=a_0+b_1X_1+b_2X_2+b_3X_3+b_4X_5+b_5X_5 \quad (1.6)$$

这里X_1至X_5均指500hPa气流形势，如X_1为春季东亚大槽所在经度，X_2为西北与华北区位势高度差等等。

5.有关问题的讨论

经验回归方程的准确与否，对农业气象各方面的工作，都是有重要意义的，因而已引起研究者的重视。本章的叙述，系围绕着用母数据配合农业气象线性回归方程是否客观的讨论，可见用蒙特卡罗模拟检验，在目前，是解决这类问题的有效途径。当然，在用这种检验方法之前，必须在已有数据的基础上，先作出一个主观方程。由于检验的结果很可能是不显著的，因此，需要作多次检验，这样多次重复，当然是件很麻烦的事。但是，为了能确切地解决农业气象学中某些有意义的问题，这样做仍是很有价值的。在目前国内计算机设备条件较好的前提下，利用编好的成套程序，运行起来也是很方便的。

主观农业气象方程的制作，也并不是纯用数字乱碰。寻求备选自变量时，必须先考虑有关的生物物理意义，如我们在确立粘虫害与高空气流的关系时就是这样。大约在50至60年代，我国研究粘虫的专家曾认为北方二代粘虫不是由南方