

数学模型 在林业中的应用研究

SHUXUE MOXING

ZAI LINYE ZHONG DE YINGYONG YANJIU

胥辉·编著

云南出版集团公司
云南科技出版社

数学模型

在林业中的应用研究

胥 辉 · 编著



江苏工业学院图书馆
藏书章

云南出版集团公司
云南科技出版社

· 昆明 ·

图书在版编目 (CIP) 数据

数学模型在林业中的应用研究/胥辉编著. —昆明: 云南科技出版社, 2009. 4

ISBN 978 - 7 - 5416 - 3250 - 1

I. 数… II. 胥 III. 数学模型—应用—林业—研究
IV. S7

中国版本图书馆 CIP 数据核字 (2009) 第 065702 号

云南出版集团公司

云南科技出版社出版发行

(昆明市环城西路 609 号云南新闻出版大楼 邮政编码: 650034)

昆明理工大学印务包装有限公司印刷 全国新华书店经销

开本: 787mm × 1092mm 1/16 印张: 20 字数: 470 千字

2009 年 4 月第 1 版 2009 年 4 月第 1 次印刷

定价: 36.00 元

编委会名单

参加本书撰写其他成员(按章序排列)

主编：胥 辉

编委：吴明山 王海亮 胡 洁 刘小菊

前言

近几十年来,随着科学技术的发展和社会的进步,数学这一重要的基础学科迅速地向自然科学和社会科学的各个领域渗透,并在工程技术、经济建设及金融管理等方面发挥出愈来愈明显、甚至是举足轻重的作用。一个现实世界中的问题,包括科学技术中的问题,往往并不是自然地以一个现成的数学问题表现,而是需要通过各种数学工具进行抽象、归结才能得出该问题的数学关系。进而利用数学的概念、方法和理论进行深入的分析 and 研究,从而从定量或定性的角度,为解决现实问题提供精确的数据或可靠的指导。一直以来,人们不断地利用数学工具解决各种各样实际问题的过程中,通过大量归结数学模型的实践,已逐步发现和总结了一些建立数学模型的规律,数学模型这一新的学科分支便因此应运而生。

以解决某个现实问题为目的,从该问题中抽象、归结出来的数学问题就称为数学模型。较著名的数学模型的定义是本德(E.A.Bender)给出的,他认为,数学模型是关于部分现实世界为一定目的而作的抽象、简化的数学结构。更简洁地,也可以认为数学模型是用数学术语对部分现实世界的描述。

利用数学模型描述客观世界是科学研究和技术开发的一种重要手段。这种方法在力学、物理学和工程技术中早已获得巨大的成功,在化学中也取得了重大进展。数学曾经被认为在生物学中的作用等于零,可是到了20世纪,数学在生物学中也起到了越来越大的作用。数量遗传学和数量生态学等新兴学科相继诞生,标志着数学这个古老学科已经成为生物学研究的一种重要工具。数学在生物学中应用的一种主要形式就是生物数学模型。在第二次世界大战以后,随着计算机技术的高速发展,数据库、数值模拟等技术已经逐渐成为许多行业不可缺少的基础。在这种环境中,到20世纪70年代以后,用数学模型方法研究生物现象,已形成一种浪潮,冲击着生物学的许多领域。

数学模型应用于生物学研究,就是利用数学语言和数学工具,描述生物学现象和规律,通过数学或逻辑推理得到一些结论,然后再将这些结论用来解释、预测生物学现象和发现新的规律。生物学现象与力学现象、物理学现象甚至化学现象有着一些非常明显的不同特点。首先,生物的特点在于它的多样性,即差异性。其差异性不仅在于物种的差异,也在于生物体个体的差异,没有两个绝对相同的生物体。其次,生物与众多的环境因素相关,决定了生物学现象的复杂性。第三,在生物与环境的协同进化中所发生的变异,导致“同一物种”与“同一环境”在不同时间可能有不同的关系。这三个特点决定了在生物学中的数学模型几乎不存在普遍适用的“常数”和所谓的“确定性模型”,甚至难以定义什么是测量对象的“真

值”。因此，深入地揭示生物学研究对象的内在规律性需要借助数学的手段，既而，数学模型成了生物学研究的重要工具之一。

林业中各研究对象同样也不会表现为现成的数学问题，必须从实际问题出发，对研究对象有关的数据进行数理统计分析，归纳出各研究因子之间的数学关系。本书即以林业中的部分数学问题为研究对象，结合林学、生态学理论，利用数学模型的方法研究其内在的规律性。本书适宜从事林业及林业数学模型相关研究的科研人员、生产实践中的林业工作者阅读，也可供林业院校林学、森林测计学等相关专业的师生作为参考书。

本书研究主要是在1999-2002年云南省自然科学基金项目“思茅松中、幼龄林生长收获模型系统与经营技术的研究”（编号：1999C0061M），2004-2009年云南省中青年学术技术带头人后备人才项目（2004py01-17），2002-2005年云南省“十五”攻关“思茅林区可持续发展关键技术研究示范”子课题“思茅松林轮栽立地生产力保持与恢复技术的研究”项目的资助下完成。本书出版得到了“云南省省院省校合作咨询、共建省级重点学科项目”资助。

本书是多位作者共同努力，集体劳动的结晶。本书各章节作者及主要研究内容如下：第一章（胥辉、吴明山）；主要研究林业中模型的共线性问题、异方差问题、度量误差问题、误差传递问题等，并使用诸如岭回归、主成分回归、偏最小二乘回归、加权回归、度量误差模型方法、联立方程组方法、二步估计等多种方法解决此类问题。第二章（胥辉、王海亮）；主要以思茅松天然次生林为对象，研究其地位指数表的编制、林分密度指数表的编制、优势木树高生长模型、树高曲线模型、林分直径结构模型、断面积生长模型、平均胸径生长模型、蓄积量生长模型。第三章（胥辉、胡洁）；主要研究思茅松土壤养分诊断问题，针对土壤中的氮、磷、钾及有机质，确定思茅松天然林、次生林、人工林的土壤养分评价指标临界值，并对思茅松三代林土壤养分进行了综合评价。第四章（胥辉、刘小菊）；主要研究思茅松轮栽后立地退化问题，使用灰色关联分析、主成分分析、数量化立地质量得分表等方法分析了思茅松立地退化程度。

本书的完成得到了西南林学院有关同仁和领导的大力支持与帮助，在此表示衷心感谢。我们也特别感谢普洱市林业局、墨江县林业局、景谷县林业局等林业系统的有关领导和同行在本书外业调查中给予的大力支持和帮助。

由于本书作者水平有限，加之书中所涉及的是一个不断变化和进步的领域，成书过程也比较仓促，书中定有不少缺点错误，敬请各位读者批评指正！

胥 辉

目 录

第一章 林业中预估模型的估计方法应用研究

1 绪论	(3)
1.1 前言	(3)
1.2 研究概述	(4)
1.3 国外研究状况	(11)
1.4 国内研究状况	(12)
1.5 研究的发展趋势	(18)
2 研究的内容和方法	(19)
2.1 林业建模中共线性问题及研究方法	(19)
2.2 异方差问题及研究方法	(19)
2.3 度量误差问题及研究方法	(22)
2.4 联立方程组模型及研究方法	(23)
3 林业中模型的共线性研究	(24)
3.1 模型构造	(24)
3.2 模型中的共线性检验	(25)
3.3 共线性的处理方法	(31)
3.4 总结	(38)
4 模型中的异方差研究	(39)
4.1 模型的选择	(39)
4.2 模型中异方差的检验	(39)
4.3 异方差的消除	(43)
4.4 加权估计与普通最小二乘估计对比分析	(49)
4.5 思茅松树高曲线方程中异方差问题的进一步研究	(53)
4.6 总结	(56)
5 材积模型中的度量误差影响及参数估计研究	(58)
5.1 引言	(58)

5.2	度量误差对落叶松材积模型的影响	(58)
5.3	带有度量误差的材积模型的参数估计	(64)
5.4	总结	(69)
6	二元立木材积模型的联立方程组模型估计方法研究	(70)
6.1	二元立木材积模型	(70)
6.2	二元立木材积模型的联立方程组模型方法	(71)
6.3	对比分析	(72)
6.4	总结	(74)
7	结论、创新和讨论	(75)
7.1	结论	(75)
7.2	创新	(77)
7.3	讨论	(77)
	参考文献	(78)

第二章 思茅松天然次生林林分生长模型研究

1	绪论	(85)
1.1	前言	(85)
1.2	林分生长模型评述	(86)
1.3	思茅松生长规律研究现状	(96)
2	研究地概况	(97)
3	外业调查	(98)
3.1	资料收集	(98)
3.2	数据整理	(99)
4	各模型评价指标的建立	(100)
5	地位指数表的编制	(101)
5.1	研究现状及存在的一些问题	(101)
5.2	数据的整理	(103)
5.3	导向曲线的拟合	(104)
5.4	数据的取舍以及导向曲线的再次拟合择优	(106)
6	林分密度指数表的编制	(109)
6.1	林分密度的定义	(109)
6.2	模型的建立	(109)

7 优势木树高生长模型	(111)
7.1 定义	(111)
7.2 模型的建立	(111)
8 树高曲线模型	(113)
8.1 定义	(113)
8.2 模型的建立	(113)
9 林分直径结构模型	(115)
9.1 研究现状	(115)
9.2 研究方法	(116)
9.3 模型的建立	(118)
10 断面积生长模型	(121)
10.1 概述	(121)
10.2 模型的建立	(121)
11 平均胸径生长模型	(125)
11.1 概述	(125)
11.2 模型的建立	(125)
12 蓄积量生长模型	(129)
12.1 概述	(129)
12.2 模型的建立	(129)
13 前两种方法所得模型的检验、比较及方法的择优	(133)
13.1 优势木树高生长模型	(133)
13.2 胸高断面积生长模型	(134)
13.3 平均胸径生长模型	(135)
13.4 蓄积量生长模型	(136)
14 选出的最优方法与系统动力学法的对比	(137)
15 Visual Basic 编程	(144)
16 结论	(145)
参考文献	(145)

第三章 思茅松土壤养分诊断研究

1 绪论	(151)
1.1 前言	(151)



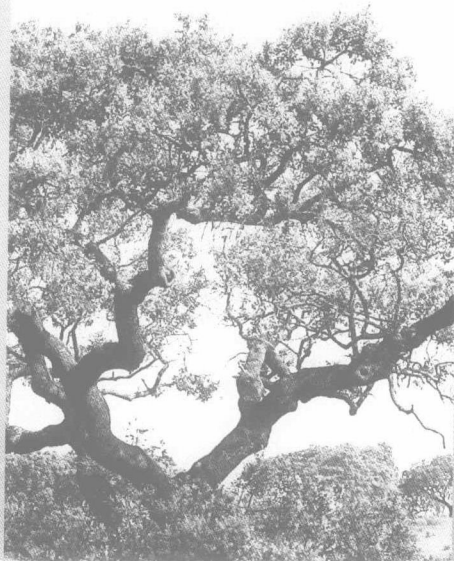
1.2	土壤养分研究概况	(152)
1.3	国外研究状况	(155)
1.4	国内研究状况	(156)
1.5	发展趋势	(158)
2	研究区概况	(159)
3	外业调查	(160)
3.1	资料收集	(160)
3.2	原始林及次生林的外业调查工作	(160)
3.3	人工林的外业调查工作	(161)
3.4	内业整理分析	(162)
4	研究方法概述	(164)
4.1	土壤养分与作物生长发育的关系	(164)
4.2	数据的标准化	(165)
4.3	快速聚类分析方法概述	(166)
4.4	主成分分析法概述	(166)
4.5	模糊综合评判法概述	(168)
5	思茅松原始林土壤养分评价指标临界值的确定	(171)
5.1	全 N 的临界值确定	(174)
5.2	全 P 的临界值确定	(176)
5.3	全 K 的临界值确定	(178)
5.4	有效 N 的临界值确定	(179)
5.5	有效 P 的临界值确定	(181)
5.6	有效 K 的临界值确定	(183)
5.7	有机质的临界值确定	(185)
5.8	小结	(187)
6	思茅松次生林土壤养分评价指标临界值的确定	(188)
6.1	全 N 的临界值确定	(191)
6.2	全 P 的临界值确定	(193)
6.3	全 K 的临界值确定	(195)
6.4	有效 N 的临界值确定	(197)
6.5	有效 P 的临界值确定	(199)
6.6	有效 K 的临界值确定	(201)
6.7	有机质的临界值确定	(203)

6.8 小结	(205)
7 思茅松人工林土壤养分评价指标临界值的确定	(206)
7.1 全 N 的临界值确定	(209)
7.2 全 P 的临界值确定	(210)
7.3 全 K 的临界值确定	(212)
7.4 有效 N 的临界值确定	(214)
7.5 有效 P 的临界值确定	(216)
7.6 有效 K 的临界值确定	(218)
7.7 有机质的临界值确定	(220)
7.8 小结	(221)
8 思茅松三代林土壤养分综合评价	(223)
8.1 思茅松三代林的各土壤养分评价指标权重的确定	(223)
8.2 思茅松三代林各土壤养分评价指标隶属度的确定	(224)
8.3 思茅松三代林土壤养分状况综合指标值的计算	(228)
8.4 思茅松三代林土壤养分综合评价结果分析	(234)
9 结论与讨论	(237)
9.1 结论	(237)
9.2 讨论	(238)
参考文献	(239)

第四章 思茅松轮栽后立地退化的研究

1 绪论	(247)
1.1 前言	(247)
1.2 研究目的	(248)
1.3 国外研究状况	(248)
1.4 国内研究状况	(249)
1.5 立地退化研究的发展趋势	(250)
2 立地退化研究概述	(252)
2.1 与立地退化有关的几个概念	(252)
2.2 立地退化研究方法的评述	(253)
3 研究区概况	(255)
3.1 思茅松简介	(255)

3.2	调查区概况	(256)
4	外业调查	(258)
4.1	资料收集	(258)
4.2	原始林的外业调查工作	(258)
4.3	人工林的外业调查	(260)
4.4	数据整理	(260)
5	研究内容和研究方法	(261)
5.1	灰色关联分析	(261)
5.2	主成分分析法	(271)
5.3	数量化立地质量得分表法	(283)
6	立地退化程度分析	(298)
6.1	模糊综合评判的基本原理	(298)
6.2	模糊评判过程	(298)
6.3	模糊评判的结果	(299)
7	小结	(300)
7.1	灰色关联分析的小结	(300)
7.2	主成分分析方法的小结	(300)
7.3	数量化理论 I 的小结	(301)
8	结论	(302)
	参考文献	(303)



SHUXUE MOXING
ZAI LINYE ZHONG DE YINGYONG YANJIU

第1章

林业中预估模型 的估计方法应用研究



1 绪论

1.1 前言

定量化是生物学科,也是林业科学发展现代化的重要标志之一。林业科学发展至今,如果只有定性的结论已不能满足实践的需要。统计分析是使林业科学结论定量化的重要工具之一。掌握统计分析方法是今天林业工作者应该具备的基本素质,这是因为与林业有关的各种现象中普遍存在着随机现象。林业是大农业的重要组成部分,森林作为最大的陆地生态系统,时空宏博、变化万千、受到许多随机因素的影响,表现为各种各样的随机现象,而统计分析正是从数量方面揭示大量随机现象中存在的必然规律的学科。

在林业工作中,从测树制表到资源清查,从种苗分级到立地质量评价,从林地中生物的分布格局到群落分类排序,从生长模型到实变预测……凡此种种都可以应用统计分析方法,至今也都积累了各种各样的成功经验。长期以来统计分析一直在各行各业直接而具体地被人们应用着,是一门在实践中应用十分活跃的数学学科分支。

我国林业科学工作者一直就非常重视统计分析这一科学研究方法与分析手段的应用。随着我国林业科学和应用数学的发展,现代林业本身也必须使林业发展的各种客观变化规律同数学有机地结合起来,并用所构造的数学模型来描述、解释和解决林业发展中所遇到的问题。用统计分析方法所建模型的质量就直接关系到林业应用中的可靠性和森林经营决策的正确性^[1]。因此,作为林业工作者,掌握正确的建模理论和统计分析方法是十分重要的,但长期以来,各种统计分析方法在林业中是否被正确运用却是一个问题。

在统计领域,对于模型估计可选的参数估计或非参数估计方法有很多,其研究成果也相当成熟。基础的有最小二乘法、广义最小二乘法、全最小二乘法、抗差最小二乘法、加权最小二乘法等;常用的还有逐步回归、稳健回归、主成分回归、岭回归、广义岭回归等。但不同的方法适用不同的条件,当遇到一些特殊的数据条件时,我们使用了不当的估计方法就会出现各种难以预计的结果。

目前,相对于统计领域,在林业统计回归建模中能被使用的统计方法比较有限,或因林业领域中的一些特殊性而造成方法的使用不当也时有出现。无论什么样的数据一般都用统计软件处理,但不知统计软件的常规算法是基于最小二乘法,如:Excel、SPSS、SAS、Matlab 等的一般线性回归和部分非线性回归。当建模的数据因子出现不满



足最小二乘法条件的时候,所得出的模型结果就变得很差,得出不符合实际意义的参数估计,从而导致不正确的结论,无法应用于林业生产中。如出现多重共线性问题时,就常用逐步回归,但逐步回归却只是一种经验的思路,理论上不能保证经过逐步回归后多重共线性一定会减低到可以接受的程度。再者,逐步回归虽然具有统计意义,但它可能削去了模型本来需要的变量信息。还有异方差问题及其加权回归中的权函数选择问题,因变量和自变量都存在误差的回归问题等等。当出现这样的情况时,如果不对数据进行改造或采用特殊的方法进行估计,必然会得出与实际结果相去甚远的参数估计。

本章就针对林业中模型估计方法的应用问题进行深入的研究,对比在克服多重共线性,消除异方差、自变量因变量同时存在度量误差时建模方面的几种估计方法,为林业建模在特定情况下提供比较恰当的参数估计方法。使林业基础研究中可选择更多的模型估计方法来改进部分模型的外推性能、参数的稳定性、模型的适用性等都具有十分重要的意义。

1.2 研究概述

1.2.1 共线性概述

在多元线性回归中,如果自变量中有两个或多个解释变量出现了较强的相关性,我们就称模型中存在多重共线性。多重共线性虽然不违背最小二乘法的四个基本假设,但会使最小二乘回归估计的结果变得很坏。

(1) 完全共线性的情况下,不存在,就无法得到参数的估计量。

(2) 近似共线性情况下,由于引起主对角线元素较大,使参数估计值的方差增大,最小二乘法的参数估计变得无效。

(3) 在共线性存在的情况下,参数并不反映各自与被解释变量之间的结构关系,而是反映它们对被解释变量的共同影响。这时,参数已经失去了应有的含义,于是经常会表现出似乎反常的现象:例如参数本来应该是正的,结果恰是负的。

(4) 模型的预测功能失效,变大的方差容易使区间预测的“区间”变大,使预测失去意义^[2]。

林业中经常会用到多自变量对因变量的线性回归问题,而在各自变量中存在线性相关性又经常出现。林业科学中诸如立木生长指标:年龄、树高、胸径、材积、枝下高、冠幅……它们之间就存在着一定的相关性^[3]。用这样的具有相关性的一些因子构建多元线性模型,难免就会出现多重共线性的可能。

在多元统计分析中,为了克服多重共线性引起的困难,人们提出了许多补救方法,

有些方法保留了完整的估计回归系数的最小二乘法,但这些估计方法中加进了限制。

(1) 众所周知,严重的多重共线性通常并不影响拟合模型用来推断应变量均值或进行预测的有效性,只要用来推断的自变量也有相同的多重共线性,它的取值与建立回归模型的数据一样。因此,一种补救方法就是限制拟合回归模型的使用,只限在服从相同多重共线性形式的自变量范围内进行推断,但在林业应用模型中,限制了模型的使用范围就失去了模型建立的意义。

(2) 为了降低多重共线性,人们经常还使用逐步回归,可以从模型中剔除一个或几个自变量,这样就可以降低模型中剩余变量的估计回归系数的标准差。但这种补救方法有几个重要的局限。第一,得不到被剔除变量的直接信息。第二,模型中剩余变量回归系数的大小受模型外的相关变量的影响。第三,这只是一种经验的思路,理论上不能保证经过逐步回归后多重共线性一定会减低到可以接受的程度。第四,在剔除后剩余变量中可能同样有强的共线性。第五,自变量的剔除是在统计学中有意义,在模型中,这些自变量有些可能不能被剔除。

(3) 有时可能增加一些能够破坏多重共线性的观测值。但是,这种方法常常是无法使用的,在林业建模中,许多自变量是不受控制的。所以,新观测值势必与以前的观测值具有相同的多重共线性形式^[4]。

1.2.1.1 岭回归消除共线性

岭回归 (ridge regression) 是 A E Hoerl 在 1962 年提出来的,其后由 R W Kennard 系统发展起来的一种改进最小二乘估计方法。

在最小二乘法计算过程中,定义

设 $0 \leq k < \infty$, 称 k 为偏常数,

$$b(k) = (x^T x + kI)^{-1} x^T y$$

就是用 $x^T x + kI$ 代替最小二乘法中的 $x^T x$, 人为降低均方误差。我们取不同的 k 值,划出 $b(k)$ 的岭迹图,作岭迹分析,求出最优 k 值,最后确定在最优 k 值下的岭回归估计 $b(k)$ 。

岭回归是通过修改最小二乘法,允许回归系数的有偏估计量存在而提出的若干种补救多重共线性问题的方法之一。如果一个估计量只有很小的偏差,但它的精度大大高于无偏估计量,人们就可能更愿意选择这个估计量,因为它更接近于真实参数值的可能性较大。而在林业建模中,常常会遇到这样的情况,所以,我们在林业建模中不能只是简单地使用一种方法,在特定的情况下我们需要使用不同的方法^[4]。

岭回归方法在统计领域中已研究得相当成熟,在各行各业中也得到广泛应用。因为岭回归的拟合数据就算有微小的变化,几乎都不影响岭回归估计值,就此意义而言,岭回归估计量是稳定的。相比之下,如果自变量高度多重共线性,最小二乘估计量可能非常不稳定。

岭回归也有其局限性。岭回归的一个很大局限就是普通的推断方法不能应用,而且精确的分布性质是未知的。另一个局限是偏倚常数 k 值的选择是凭判断的,方法不同或主体不同,其结果也不相同。正是偏倚常数 k 值难以确定,在林业统计中该方法