

第二次全国概率统计天气预报 会议论文集

会议论文集编辑小组 编

科学出版社

P45-53

223

第二次全国概率统计天气预报 会议论文集

会议论文集编辑小组 编



科学出版社

1986

32152

内 容 简 介

本文集收集了第二次全国概率统计天气预报会议上宣读的 39 篇论文。内容涉及门限自回归、非量度的多维变换、随机函数典型分解的应用、相关系数的不稳定性、逐步筛选因子的聚类分析、不规则格点上车贝雪夫多项式的应用、时空谱分析、模糊数学的应用、随机动力模式、卡曼滤波、Schrödinger 方程的应用、农业气候区划等。

本书可供气象台站天气预报工作者、气象科学研究工作者以及水文、地质、地震等部门的工作者和高等院校师生参考。

第二次全国概率统计天气预报 会议论文集

会议论文集编辑小组 编

责任编辑 许贻刚

科学出版社 出版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1986 年 10 月第 一 版	开本：787×1092 1/16
1986 年 10 月第一次印刷	印张：14
印数：0001—1,400	字数：316,000

统一书号：13031·3312

本社书号：4627·13—15

定 价：3.30 元

前 言

我国气象工作者，在三十年代即开始从事概率统计天气预报的研究，取得了一些成果。解放以后，概率统计预报有了较大的发展。1973年初，召开了全国第一次数值预报和数理统计预报会议。同年四月，在南京举办了全国概率统计预报短训班。短训班介绍了现代统计预报方法和我国气象工作者所取得的成果。这个短训班，对推动全国统计预报的发展，培养统计预报骨干起了良好的作用。1973年以后，我国统计预报的研究和实践有了巨大的发展。1981年5月中国气象学会在烟台召开的全国概率统计天气预报会议，检阅了近年来我国概率统计预报研究和应用方面的优秀成果，研究了进一步开展统计预报科研协作的问题。会上宣读的论文表明，我国不仅在概率统计预报理论的各个领域取得了很多成果，而且在将概率统计方法应用于台站业务预报实践方面有了广泛的发展。

会议的主要学术成果反映在如下几个方面：

(1) 多元分析：引入了数量化方法、非线性映象、图象识别、模糊数学、非量度的多维变换、对应分析等各种方法。相似预报、聚类分析等在台站普遍应用，同时还提出了动态相似问题。讨论了各种距离之间、数量化与回归、判别之间的关系及相关系数不稳定性等问题。自然正交分解的应用更加广泛。同时还提出了不规则格点上的车贝雪夫多项式展开问题，并将其用于要素场的分析和预报之中。模糊数学方法已用于降水天气周期、春播低温阴雨、寒潮、台风路径的分析和预报等方面，并取得了初步的成果。

(2) 时间序列：引入了 ARIMA 模式，并用之于季节预报。门限自回归模式也得到了初步应用。证明了自回归模式的最后几个系数就是偏相关系数，因此可用 F 检验来截尾。利用波谱分析方法揭示了超长波活动，并引进了最大熵谱、时空谱等较新的方法。周期均值叠加法是我国气象工作者提出的，现已在理论上证明了它的收敛性。

(3) 统计动力预报：继气象台实现了模式输出之后，又在气象站应用模式输出预报台风、暴雨方面取得了成果。研究了多起始场的一阶随机差分模式，并在实际预报中做了试验。讨论了卡曼滤波方法用于台风路径预报的问题，提出了利用量子力学观点研究长期天气过程的问题。在关于随机动力气候学的研究中，考虑了大气与海洋之间的相互作用。

(4) 概率统计的理论与方法：研究了等概率预报集成的问题，预报因子与预报准确率的关系问题等，还涉及了关于预报模式的检验问题、极值预报问题及非正态模式问题。

(5) 气候统计与台站预报：各种统计方法已用于气候分析、农业区划等各个方面。概率统计预报已用于晴雨、降水、暴雨、雷暴、春播连阴雨、冰雹、寒露风、台风、寒潮等的长中短期预报之中。

本文集收集了烟台会议上宣读的大部分优秀论文，这些论文涉及气象统计预报、天气气候分析等各个方面。

烟台会议上，由会议学术组成员史久恩、朱洪绩、朱盛明、沈长泗、李黄、**李麦村**、周家斌、范钟秀、姚棣荣、章少卿、梁幼林、陈国范、洪仁宗、曹鸿兴、黄忠恕、黄嘉佑、谭冠日、

滕家谟集体讨论,提出初选论文,并委托周家斌、黄嘉佑、曹鸿兴具体组织稿件,邀请汪关成为技术编辑。初选论文由学术组在京成员审阅,最后审定 39 篇论文编入文集。

由于我们学术水平有限,又缺乏经验,因而本文集中还可能存在不少问题,敬请读者批评指正。

中国气象学会第二次全国概率统计

天气预报会议学术组

目 录

气候概率统计学在统计预报与统计气候学方面的进展·····	么枕生 (1)
门限自回归模型在天气预报中的应用·····	项静恬 黄文杰 (6)
双重筛选多元逐步回归在物理统计预报中的应用·····	沈元芳 (14)
回归模型在气象预报中的应用·····	汤康恩 (21)
运用天气学原理建立长期数理统计预报模式·····	储锁龙 谢 庄 (24)
北京地区雷暴因子两种 0, 1 化方法的探讨及预报方程分析·····	戴淑芬 何于班 (29)
6 月份冰雹的概率回归预报·····	李文源 (33)
用数理统计方法作 7 月份低槽型雷暴的短期预报·····	韩广志 吴建中 (36)
多元分析中相关稳定性问题·····	王立生 (39)
用数量化理论 (I) 作台风路径预报·····	姚棣荣 金一鸣 (43)
数量化方法 (I) 和聚类分析的若干应用问题·····	郑祖光 (49)
聚类分析在气象上应用的几个问题·····	滕家谟 (56)
逐步筛选因子的聚类分析法及其应用·····	郭永润 季明珠 (59)
逐步聚类分析及其在客观分类预报中的应用·····	金一鸣 (64)
近百年全球海平面气压 1, 7 月距平场的聚类分析 ·····	史久恩 马怀存 周琴芳 王隽骧 (69)
用筛选因子最优分割法进行气候-农业产量区划·····	丁士晟 (75)
关于汛期月雨量分布型式的预报问题·····	王得民 (80)
气象要素水平分布的统计预报方法(二)——我国东北地区夏季气温·····	
分布的预报·····	周家斌 (86)
非量度多维变换在气象上的应用·····	杨保桂 (93)
随机函数典型分解及其在水文长期预报中的应用·····	范钟秀 (99)
赤道太平洋海温场与太平洋副热带高压的关系的统计分析 ·····	黄嘉佑 李麦村 (103)
大尺度运动响应过程的若干特征及其动态谱估计·····	李麦村 梁幼林 (110)
春季环流的时-空谱分析·····	李 黄 刘桂芝 (118)
用互谱与距离函数研究三年周期现象·····	马镜娴 (125)
功率谱和方差分析在年轮气候分析中的应用·····	王玉玺 (130)
近百年山西省旱涝的周期性·····	刘幼唐 (133)
单要素长期预报的时间序列方案·····	罗乔林 吕致君 (136)
横向时间序列分析和预报检验·····	汤仲鑫 (144)
一个考虑单站多要素连续演变的统计预报方法·····	盛家荣 (147)
用时间序列预报 7—9 月西北太平洋台风个数·····	邱志生 (150)
模糊集原理在长期天气预报中的应用·····	陈国范 曹鸿兴 (154)

模糊推理及其在预报评分中的简单应用.....	孔佑坤 (162)
海气相互作用的随机气候模型	李麦村 黄嘉佑 (169)
正压随机差分模式的初步计算结果.....	朱盛明 曹鸿兴 (175)
台风路径预报的卡曼滤波方法.....	李贤琅 (183)
一种动态相似预报方法的讨论.....	章少卿 (189)
月、季平均环流的几率波特性及 Schrödinger 方程的应用	李麦村 章少卿 (192)
条件概率与预报指标的若干讨论.....	马天龙 (202)
天气预报的概率评价.....	沈长泗 (208)

THE COLLECTION OF PAPERS PRESENTED AT THE SECOND CONFERENCE ON PROBABILITY AND STATISTICAL METHODS IN WEATHER FORECASTING

CONTENTS

Foreword	
Development of climatic statistics in statistical forecasting and statistical climatology	Yao Zhensheng (1)
Application of threshold autoregressive model to weather forecasting	Xiang Jingtian, Huan Wenjie (6)
Application of double screening stepwise regression to physicostatistical weather prediction	Shen Yuanfang (14)
Application of regressive model to meteorological prediction	Tang Kang'en (21)
A statistics long-range forecasting model based on synoptic principles	Chu Suolong, Xie Zhang (24)
Investigation of two different binary-digital methods for screening predictors for thunderstorm forecasting and analysis of forecasting equation	Dai Shufen, He Yuban (29)
Prediction of hailstorm in June using REEP	Li Wen yuan (33)
Short-range forecast of thunderstorm in July based on statistics	Han Guanzhi, Wu Jianzhong (36)
On correlation stability in multivariate analysis	Wang Lisheng (39)
Predicting typhoon track with quantitative method (I)	Yao Dirong, Jing Yiming (43)
Problems in application of quantitative method (I) and cluster analysis	Zhen Zuguang (49)
Problems in the application of clustering to meteorology	Teng Jiamo (56)
Cluster analysis of stepwise screen predictors and its application	Guoy Yongrun, Ji Mingzhu (59)
Application of stepwise cluster analysis to objective classification prediction	Jing Yiming (64)
Cluster analysis on monthly anomaly of global sea-level pressure in recent hundred years	Shi Jiuen et al. (69)
A method of optimal division of screening factors as applied to climate-agriculture product division	Ding Shisheng (75)
On the prediction of the distribution pattern of monthly precipitation in the rainy season	Wang Demin (80)
Statistical forecasting method for horizontal distribution of meteorological elements (II) —Temperature distribution forecasting over the Northeast China in summer	

.....	Zhou Jiabin	(86)
Application of nonmetric multidimensional scaling to meteorology.....	Yang Baogui	(93)
Canonical decomposition of stochastic function and its application to medium-long range hydrological forecasting	Fan Zhongxiu	(99)
Statistical analysis about the relationship between sea surface temperature in equatorial region and subtropical high pressure system	Huang Jiayou, <u>Li Maicun</u>	(103)
Some characteristics of the response processes of large scale motion and its dynamic spectral estimation	<u>Li Maicun</u> , Liang Youlin	(110)
Space-time spectral analysis for the general circulation in spring.....	Li Huang, <u>Liu Guizhi</u>	(118)
A study on 3-year period with cross-spectrum and distance function.....	Ma Jingxian	(125)
Application of power spectrum and variance analysis to tree-ring climatic analysis	 Wang Yuxi	(130)
The periodicity of drought and flood in Shanxi province in the recent 100 years	 Liu Youtang	(133)
Time series analysis of long-range forecasting of meteorological elements	 Luo Qiaolin, Lü Zhijun	(136)
Transverse time series analysis and forecasting verification.....	Tang Zhongxin	(144)
Statistical forecast based on multi-element sequence of single station	 Sheng Jiarong	(147)
Prediction of the number of typhoon in July to September over north-west Pacific using time series analysis.....	Qiu Zhisheng	(150)
Application of fuzzy set theory to long-range weather forecasting	Chen Guofan, Cao Hongxing	(154)
Fuzzy reasoning and its simple application to forecasting score	 Kong Youkun	(162)
Stochastic climatical model of sea-air interaction.....	<u>Li Maicun</u> , Huang Jiayou	(169)
The preliminary study of barotropic stochastic difference model	 Zhu Shengming, Cao Hongxing	(175)
Kalman filtering method for typhoon track prediction	Li Xianrang	(183)
On one dynamic analogue forecasting method.....	Zhang Shaoqing	(189)
Characteristic of probability wave of monthly-seasonal mean circulation and application of Schrödinger equation.....	<u>Li Maicun</u> , Zhang Shaoqing	(192)
On conditional probability and forecasting index	Ma Tianlong	(202)
Probability evaluation of weather forecasting.....	Sheng Changsi	(208)

气候概率统计学在统计预报与统计 气候学方面的进展

么 枕 生

(南京大学气象系)

一、引 言

作者多年来早已认为气候学内容不仅只有气候分析,也应有气候预报。所以气候概率统计学的进展方向,有以统计分析为主的统计气候学方向,还应有统计预报的方向。

作者也认为天气预报与气候预报不同之点,不是时间尺度的长短不同,而是预报对象性质的不同。不论预报时间长短,能够预报天气过程的预报都是天气预报,只能预报气候统计量或天气气候过程的预报都是气候预报。从这一概念来看,在气象学领域内除用相似原理做长期的天气预报外,许多统计预报的内容应为气候预报。现在作者就用这个观点,总结气候概率统计学在统计预报方面的进展,并提出建立统计气候学的问题。如果作者的观点是正确的,那么统计预报和统计气候学必然都是在气候概率统计学的基础上发展的,并且是彼此相互联系的两个发展方向。

二、统计预报的进展

气候学研究的对象是记录序列,统计方法自然是很重要的工具,所以气候学家 J. Hahn (1902) 早就讨论和检验过序列中的均匀性以及序列的订正问题。Hahn (1926) 曾用误差理论检验气候要素变量是否符合正态分布,并首次 (1917) 提出研究谐波分析的重要性。这个统计方法是程纯枢 (1942) 首次引进到中国的。Fourier 早已证明复杂周期观象乃是由许多周期、振幅、相角不同的谐波组成的。Strachey (1887) 还提出谐波系数的表达公式。现行的谐波分析方法是 Pollak (1920) 提出的。Haurwitz 与 Craig (1952) 提出了球面谐波, Kubote 等 (1960) 进而用球面谐波做正压预报。

现在沿用的周期分析法是 Schuster (1898) 提出的。中田 (1949) 提出用差分法与移动(滑动)平均法就可以分开长周期与短周期,然后分别外推,再进行合成,就可以用于预报。Brooks 与 Carruthers (1953) 首次给出了用方差分析法分析周期的示例。我国的不少台站则进一步将它发展为周期叠加外推的预报法。

另一个周期分析方法就是用自相关绘制相关图,这是 Wold (1938) 命名的。用自相关系数去确定周期,自本世纪以来已有很多文献,但 Fuhrich (1933, 1934) 的工作已使这个方法大大前进一步。他证明:如连续计算自相关系数就能使各种周期按其重要性显示

出来,且提出判别周期存在的准则。在周期分析的基础上又发展出谱分析。周期图是离散谱。谱分析方法在气象中的应用是 Panofsky (1955) 引进的。Leese 与 Epstein (1963) 又引进了二维谱分析。缪锦海与曹鸿兴分别在1979年把最大熵谱分析引进到中国。这个方法在短资料的情况下,对于低频具有较好的分辨能力,而长波确定正是统计预报所要解决的问题。最近在国际上又有谱分析方面新的方法,如自回归谱、对数功率谱、多元谱、时空谱以及谱分析的非线性方法。

Yule (1927) 对于周期的概念有重大的发展与创见,他由谐波模式推导出自回归模式。现在由 Box 与 Jenkins (1970) 发展成为各种随机模式。Davis 与 Rappoport (1974) 曾用自回归移动平均模式,预报美国的干旱,并和 Brown (1962) 的修匀模式做了比较。Jones (1964) 为了年时间序列的预报,曾讨论了估计非确定分量的谱密度问题,提出用周期图估计谱估计自回归模式中的参数。Burg (1967) 从另一观点结合最大熵谱的计算,提出估计自回归模式中参数的另一方法。

在时间序列分析中常常把不必要的频率去掉,这个过程称为滤波。我们在台站上一级常用的都是等权移动平均滤波。此外还有正态滤波与指数滤波。我们还可以设计具有特定频率响应的滤波。罗哲贤 (1977) 认为利用移动平均的滤波技术,提取低频振动时误差很大,影响长期预报的结果。所以根据脉冲响应与频率响应设计了两个低通滤波器,以便改进移动平均。王宗皓、李麦村等 (1975) 着重介绍了用卡曼滤波原理所提出的适应性卡曼更新理论,用作多元非平稳时间序列的过滤。

作者于1978年曾提出历史旱涝记载的统计方法,把史书有关旱涝的文字描述化为二分变量去做周期分析。这些方法曾由么枕生等 (1979) 用于研究我国近世纪的历史旱涝变化。用这种方法所分析的周期和用气候记录所做的周期分析一样。在用周期叠加时虽可拟合历史记录,但很难用于长期预报。在周期叠加的谐波模式中,每种周期的三个谐波参数仍是并不独立的随机变量,错综复杂,难以处理,只可根据前期征兆凭预报员的经验近似估计以后变化的趋势。

回归与相关是相联系的,起源很早。回归起源于遗传理论中研究父子身高间的相关,回归一名是人类学家 Galton 早在19世纪命名的。Hildebransson (1897) 首次提出天气气候条件的相关,用每月气压距平研究出大气活动中心。以后 Walker (1923—1938) 在普遍找相关与自相关的基础上,发现在全世界海平面气压中具有三大涛动,最后在这个基础上用回归制做长期预报。涂长望 (1937) 应用 Walker 的方法预报我国夏季的洪水与干旱。相关与回归仍为长期预报中所应用的简单客观分型与预报方法,不只我国如此,国外也是如此。Paegle 与 Keirulff (1974) 的工作就是最好例证。

李麦村 (1975) 曾全面回顾过统计预报的进展,首先指出用逐步回归筛选因子仍为普遍乐用的方法。这个统计预报方法是史久恩等 (1964) 首次引进我国的。王宗皓、李麦村等 (1975) 介绍了逐步正交回归 (正交筛选),以及根据决策论与相似原理做预报的方法。决策论也是气候概率统计学的内容。决策论在两类预报中较为成功。这个方法应用的示例是在1974年由中国科学院大气物理研究所概率统计预报组给出的。

作者 (1977) 提出用正交相关筛选建立正交回归的方法,其计算结果与性能完全和逐步回归一样,但计算量大大减少。作者 (1977, 1980) 也提出用偏相关筛选与多重相关筛选建立更为简化的多重回归方法,其计算结果与性能和逐步回归一样,只是没有正交回归

的性能。

在多元统计分析方面有车贝雪夫多项式与自然正交函数展开场量选取因子去做预报的方法。车贝雪夫正交多项式与自然正交函数是 Wadsworth 等(1948)与 Lorenz (1956)第一次应用于场量的展开去做预报的。在我国,张家诚等(1963)首次引用车贝雪夫多项式。施永年(1965)曾把车贝雪夫多项式用一维推导过。周家斌最近研究如何在不规则格点上展开车贝雪夫多项式的问题。这是发展了车贝雪夫多项式。用自然正交分解去做预报的方法是李麦村等(1971)在我国推行的。

Miller (1962) 提出用判别分析做预报的方法。这个方法是李麦村首次引进到我国的。杨自强(1976)与李麦村(1977)相继发展成为多级逐步判别方法。王得民(1978)曾给出详细计算示例。

典型相关分析是 Glahn (1968) 引进气象学领域的, 近来史久恩等也把这个方法引进到我国。多重线性回归仅仅是典型相关的特殊情况, 但典型相关变量计算的预报方程只能是二分变量。典型相关分析和判别分析相似, 前者要计算一个典型函数, 后者要计算一个判别函数, 不过判别分析要计算一个判别标准, 而典型相关分析却直接计算二分变量的预报量。

Gabriel (1959—1962) 曾首次用一阶马尔科夫链研究每月降水的出现。此后文献已经很多, 并且指出晴日与雨日的序列所要求的马尔科夫链阶数并不一致。还有计算条件概率的问题。用什么阶的马尔科夫链才能代表记录序列, Lowry 等(1968)曾介绍了检验方法。作者(1966)在划分状态时就考虑了历史演变, 间接作到了拟合高阶马尔科夫链。周家斌(1975)改进了计算, 结果说明这样的马尔科夫链能反映晴雨变化的全部历史演变过程。

最近赵溱(1978)提出按变量的重要程度分类, 逐步筛选找出最相似的个例去预报。陈国范与曹鸿兴(1979—1980)合作引进了模糊集理论, 建立模糊相似矩阵, 结合预报员的经验更客观地找出相似的平均形势图, 作为预报的根据。对应分析方法等价于列联表以及典型相关分析, 把 R 与 Q 型分析联系起来, 成为因子分析方法中最为综合的工具, 可以按不同的天气气候形势, 用不同的因子做统计预报。

统计预报是客观的长期预报方法, 但是最好是因子筛选和建立预报方程要直接结合起来。所以逐步回归以及自回归模式至今仍有其生命力。我们知道, 用于气候统计分析的统计模式和用于预报的统计模式应有区别, 前者注意历史记录的拟合, 后者要注意预报效果。预报效果好的统计模式往往不是拟合历史记录好的模式。预报模式要注意最近记录的拟合。Brown (1962) 提出的指数修匀模式, 就适合预报模式的一定要求。Gilchrist (1967) 提出用折扣最小二乘法拟合回归, 也有预报上的优点。这个方法是把均方误差乘以折扣因数 a^r ($0 < a < 1$) 使其成为最小时求得正规方程。 r 代表预报时段, a 愈小, 则回归线愈接近于最近记录。此外, 还有 Holt-Winter 方法以及最近发展的不同型的非线性模式。

总之, 在国际上统计预报进展很快。在推动我国统计预报的进展历程上, 中国科学院大气物理研究所和国家气象局在有关单位和高等院校的配合下, 已经做出很大的贡献。从历史上看, 这些贡献显然是从处理气候资料序列的气候统计进一步发展的结果。

三、统计气候学的进展

作者认为气候学的进展要走向边缘科学。在目前来看有五个边缘方向,即物理气候学、天气动力气候学、统计气候学、统计动力气候学与地理气候学。这就是说:以什么为研究工具的气候学,必然发展为那种研究工具的边缘气候学。根据我的观点,统计气候学的建立是必然的趋势。

作者自1941年后,就注意在教学中引进统计方法,在1951—1959年的论著都是统计气候学的内容。在统计气候学的研究中深感统计理论是重要工具。Leith (1973)认为气候就是用一个无限期间的理想记录所规定的大气规律。这就是从统计学观点给气候学所下的定义。根据这个定义所研究的气候学当然就是统计气候学,果然国际上现已建立了这门科学。气候的成因只有用物理气候学等另外四门气候学分支才能研究清楚。气候现象主要用统计气候学进行分析。例如,Craddock (1956)曾用中欧与北欧42个台站的长年候平均记录,做谐波分析的工作就是一例。他认为各地的两项谐波所以具有差异是存在一个系统性分量。在任何一个台站上的距平,不但具有序列相关,而且具有空间相关。他做谐波分析的目的,就是为了证明年温变化过程中确有气象谚语中有关定期冷暖的显著奇异性。他发现的空间相关也证明了作者(1951)用年温谐波分析划分的中国气候确有客观根据。

我们为了做好统计预报,事实上必然已经做了大量的统计气候学工作。例如,在统计预报中的相关分析、回归分析、谱分析、判别分析、典型相关分析、经验正交函数、车贝雪夫多项式、聚类分析等方法用于统计预报以前,都已经做了大量的统计气候学工作,但绝大多数只是为了做好统计预报,很少注意突出基本的统计气候学成果。

国外在统计气候学的方向上,已经提出很多工作。例如Horn和Bryson, Lettau与White,以及McGee与Hastenrath分别在1960, 1964与1966年都用谐波分析做过降水分型。这就是用统计模式研究具有时空变化的统计气候学典例。虽然真正周期是很少的,用球面谐波代表大气流型的处理方法受到很大的限制,但是这种处理方法作为初步的环流模式研究还是可以的。Elsaesser (1966)也把北半球气象记录展开为球面谐波级数。

Polowchak与Panofsky (1968)认为每月温度的谱可用做气候指标。Rayner (1965, 1967)对时间序列和空间序列应用谱分析与交叉谱分析去说明新西兰的大尺度时空气候特征。

Cchak (1962)推导二维车贝雪夫多项式的计算手续,不但计算方便,且详细说明了正交多项式展开气压场的意义与应用。这是在统计气候学领域内,用正交多项式描述天气气候形势最好的范例。Ward与Shapiro (1961)曾在北半球对经向与纬向指数、持续量数以及北美与欧洲地面气压场、西半球500毫巴高度场的多项式系数都做过谱分析,其结果说明具有持续性(红噪声)和每月与每年的循环以及两年振荡。

经验正交函数是从记录本身推导出一组正交函数,具有随机适应记录本身的特点,所以有很多作者把特征向量分析应用于气候学中的不同问题。Stidd (1967)用特征向量研究美国内华达州各地的降水气候。更有许多作者研究北半球或一个大陆尺度的统计气候学问题,例如Craddock与Flood (1969), Craddock与Flintoff (1970)就做了研究气压

高度场与厚度场的统计气候工作。

多重回归中偏回归系数是分析气候的依据,尤其适合于统计气候的研究——即所谓趋势面分析。在这样的多重回归中,应变量要用多项式拟合,因此可以把那些影响研究对象的最简单大尺度型式,从局地的随机变化中抽出来。这个方法适合于降水气候分析。

在统计气候中最简单的统计方法为相关法,它不仅从时间序列中而且还从空间序列中统计相关性。这个方法对于研究我国南涝北旱的面积与概率具有参考价值。

总之,我们在统计预报中已经做了很多统计气候学的基本工作,如何把这一基础工作更有系统地突出出来,是统计预报进展方向上的一个重要方面。

四、结 论

气象记录是确定过程和随机噪声相结合的极不规则的时间序列。我们用动力模式只能描绘气象记录中的确定部分,只能给出一个作为参考的轮廓。

在气候变化中,不只气候记录是噪声记录,更有海、地、气相互作用的问题。李麦村最近注意到这个复杂问题,提出一个长期预报随机动力模式。

长期预报是有关农业现代化和科学技术现代化所极需研究的问题。如何壮大组织,壮大阵容,帮助国内更多的同志,用集体的力量在这个方向上更攀登一步,是我国在四化号召下应当考虑的问题。

参 考 文 献

- [1] Convad, V. and L. W. Pollak, Methods in Climatology, Cambridge Mass., 1950.
- [2] Brooks, C. E. P. and N. Carruthers, Handbook of Statistical Methods in Meteorology, London, 1953.
- [3] 么枕生,气候统计,科学出版社, 1953.
- [4] 李麦村,统计预报的进展,近代气象学若干问题的进展,科学出版社, 1975.
- [5] Gilchrist, W. Statistical Forecasting, J. Wiley and Sons, 1978.

门限自回归模型在天气预报中的应用

项 静 恬

黄 文 杰

(中国科学院应用数学研究所) (国家气象局气象科学研究院天气气候研究所)

Box 和 Jenkins^[1] 在 1970 年给出的自回归求和滑动平均 (ARIMA) 模型为处理线性时间序列提供了有力的数学工具,但实际工作中所遇到的数据序列往往规律复杂,不能用线性模型来描述。因此,非线性时间序列模型的研究和应用在近几年越来越为人们所关注。本文应用的门限自回归模型 (TAR)^[2]就是一种非线性时间序列模型。

门限自回归模型的基本思想就是用逐段线性化手段来处理非线性系统;此外,门限的控制作用保证了系统的稳定性。所以,它可以有效地描述复杂的周期序列。

一、门限自回归模型的几个类型

门限自回归模型的建立依赖于门限的选择,对于一个数据序列 $\{x_i\}$,首先将它作从小到大的按顺序排列,再根据一定标准对数据序列作区间划分。为此,设 $r_0, r_1, \dots, r_l$ 为从小到大顺序排列的有限个实数(其中 $r_0 = -\infty, r_l = \infty$),它们确定了对于实数轴的一种划分:

$$R = R_1 \cup R_2 \cdots \cup R_l, \quad R_i = (r_{i-1}, r_i],$$

则 R_i 即为数据序列的第 i 个区间,该区间的端点值 r_{i-1} 和 r_i 即称为第 $i-1$ 和第 i 个门限。此外,为了最优地描述系统的规律,在建立模型时需要选取延迟步长 d 。

1. 自激励门限自回归模型

对于一个数据序列 $\{x_i\}$,可以建立如下的自激励门限自回归模型:

$$x_n = a_0^{(j)} + \sum_{i=1}^{k_j} a_i^{(j)} x_{n-i} + \varepsilon_n^{(j)} \quad x_{n-d} \in R_j \quad j = 1, 2, \dots, l \quad (1)$$

式中 k_j 为第 j 个模型的阶, $a_i (i = 0, 1, \dots, k_j)$ 称为自激励门限自回归系数, ε_n 为白噪声, l 为模型个数。

由于模型中输入与输出以及控制条件全部由时间序列 $\{x_n\}$ 本身提供,因此称之为是阶 $\{l; k_1, k_2, \dots, k_l\}$ 的自激励门限自回归模型,记作 SETAR $(l; k_1, k_2, \dots, k_l)$ 。

在特殊情况下,SETAR $(1; k)$ 正好是 k 阶自回归模型 AR(k)。

2. 开环门限自回归模型

在实际应用中常常遇到某一时间序列 $\{x_i\}$,除其本身前后间存在相关外,还与另一时间序列 $\{y_i\}$ 有着密切的关系,对于这类情形可以建立开环门限自回归模型,即

$$x_n = a_0^{(j)} + \sum_{i=1}^{m_j} a_i^{(j)} x_{n-i} + \sum_{i=1}^{s_j} b_i^{(j)} y_{n-i} + \varepsilon_n^{(j)} y_{n-d} \in R_j \quad j = 1, 2, \dots, l \quad (2)$$

式中 m_j 和 s_j 为第 j 个开环门限自回归的阶, $a_i (i = 0, 1, \dots, m_j)$ 和 $b_i (i = 1, 2, \dots, s_j)$ 为开环门限自回归系数, ε_n 为白噪声, l 仍为模型个数。

由于模型以 $\{y_i\}$ 为可观测输入, 以 $\{x_i\}$ 为可观测输出, 所以模型描述了一个 (x_n, y_n) 开环门限自回归系统, 简记为 TARSO $(l; (m_1, s_1), (m_2, s_2), \dots, (m_l, s_l))$ 。

3. 闭环门限自回归模型

若 (x_n, y_n) 和 (y_n, x_n) 都为 TARSO, 且所有白噪声序列都满足相互独立等条件, 则称此系统为满足闭环门限自回归模型, 简记为 TARSC $(l; (m_1, s_1), (m_2, s_2), \dots, (m_l, s_l))$ 。

二、门限自回归模型的性质

1. 非线性模型的逐段线性化

考虑一个一般形式的一阶非线性自回归模型, 简记为 NLAR (1):

$$x_n = f(x_{n-1}) + \varepsilon \quad (3)$$

可以证明, 当 f 满足一定条件时, NLAR (1) 可用 TAR 模型任意逼近。

由维尔斯特拉斯 (Weierstrass) 定理知, 闭区间 $[x', x'']$ 上连续函数 $f(x)$ 可由如下逐段线性的 $\hat{f}(x)$ 任意逼近:

$$\hat{f}(x) = f(x_{i_{k-1}}) + \alpha_k x \quad \text{当 } x \in [x_{i_{k-1}}, x_{i_k}] \quad k = 1, 2, \dots, l$$

而 $x_{i_0} = x', x_{i_l} = x''$, 这就给出了闭区间 $[x', x'']$ 上的一种划分:

$$[x', x''] = [x', x_{i_1}) \cup [x_{i_1}, x_{i_2}) \cdots \cup [x_{i_{l-1}}, x'']$$

划分的粗细由要求的精度决定。由此得到一个以 $x_{i_1}, x_{i_2}, \dots, x_{i_{l-1}}$ 为门限的 SETAR $(l; 1, \dots, 1)$ 来近似 NLAR (1)。

同理, 一般的 NLAR(k) 模型也可通过一定的手段达到由 TAR 逼近的目的。

2. 门限控制保证了递推的稳定性

所谓稳定递推就是要求递推关系

$$x_n = f(x_{n-1}, \dots, x_{n-k}) \quad (4)$$

能保证对于所有的 $j < \infty$ 使输出 $|x_j| < \infty$ 。为了保证系统的稳定性, 常需对函数 f 附加一定的条件, 或是对 f 的输出(值域)加以适当的限制。TAR 模型就是采用了后一种手段来保证递推的稳定。

实例和分析可以证明, 尽管模型中局部系统并不稳定, 但门限控制可以保证整体递推的稳定进行。并且由于门限的控制作用, 系统呈现有规律的向零收敛或发散, 形成一个周期性循环。

三、建模步骤和统计识别

一般非线性系统只可能在一定限制条件下建立特定的非线性模型,而门限自回归的建模可以借助线性模型的建模手段而不存在本质的困难,这也是门限自回归模型的一个优点。

本文仅就 SETAR 模型介绍建模和统计识别的具体步骤如下,其余情形可以类推。

SETAR 的建模,要求估计以下指标和参数:

- (1) 模型个数 l ;
- (2) 延迟步长 d ;
- (3) 门限值 r ;
- (4) 自回归阶数 k_j 和相应自回归系数 $a_i^{(j)} (i = 0, \dots, k_j), j = 1, \dots, l$ 。

模型的自回归系数以用线性最小二乘估计最为方便。一般说,延迟步长 d , 模型个数 l 和门限值 r 以及自回归阶数 k_j 宜根据线性估计中的最小信息量准则——AIC 准则^[3]逐层选取确定。现分述于后。

1. 先固定模型个数 l 和延迟步长 d

无论是模型个数 l (亦即对数据序列作 l 段分割) 和延迟步长 d 皆可根据物理分析或经验确定,在无法做物理分析或无经验作根据情况下,可采用统计识别办法。

2. 门限值 r_i 的确定

在 l 的值确定后,门限值 r_i 的选定是个重要的步骤,因为门限自回归模型的好坏与各段符合线性化的程度有关。

在以物理分析或经验为根据确定 l 的同时很容易确定门限值 r_i ,但若以统计识别方法确定则需要多种尝试。例如取 $l = 2$ 时,由于 $r_0 = -\infty, r_2 = \infty$, 仅 r_1 待定,这时我们可取门限值 $r_\alpha (\alpha = 0.3, 0.4, 0.5, 0.6, 0.7)$ 作尝试,以 AIC 准则确定的方法在后面给出。

3. 计算 AIC 值

给定模型最大阶数 L , 并对 l 个模型各用最小二乘估计求出自回归系数估计值和残差平方和值,并计算相应的 AIC 值。

列出各模型的回归方程

根据延迟步长 d 和门限值对原序列分别列出各个模型应参加回归的方程,例如原序列中 $x_{k_1-d+1}, x_{k_1-d+4}, x_{k_1-d+5}, \dots$ 都是在门限 r_0 和 r_1 之间时(即 $x_{k_1-d+1}, x_{k_1-d+4}, x_{k_1-d+5}, \dots, \in R_1$), 系统的第一个模型应有

$$\begin{bmatrix} x_{k_1+1} \\ x_{k_1+4} \\ x_{k_1+5} \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 & x_{k_1} & x_{k_1-1} & \cdots & x_1 \\ 1 & x_{k_1+3} & x_{k_1+2} & \cdots & x_4 \\ 1 & x_{k_1+4} & x_{k_1+3} & \cdots & x_5 \\ \vdots & \vdots & \vdots & \ddots & \vdots \end{bmatrix} \begin{bmatrix} a_0^{(1)} \\ a_1^{(1)} \\ \vdots \\ a_{k_1}^{(1)} \end{bmatrix} + \begin{bmatrix} \varepsilon_1^{(1)} \\ \varepsilon_2^{(1)} \\ \vdots \\ \varepsilon_{N_1}^{(1)} \end{bmatrix} \quad (5)$$