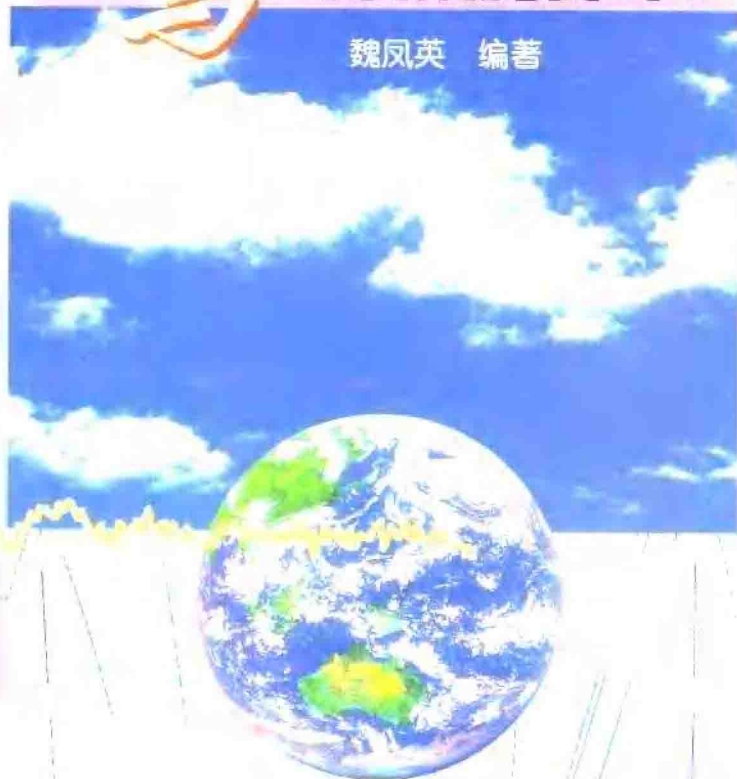


现代气候统计诊断

预测技术

魏凤英 编著



气象出版社

现代气候统计诊断与预测技术

魏凤英 编著

气象出版社

序

自本世纪 80 年代以来,随着计算机技术的迅猛发展,现代统计突破对观测资料统计指标的简单计算和分析的局限,展现了重大飞跃。这种飞跃不仅体现在统计技术内容的深度和广度上,更体现在研究问题的新思维和新观点上。与此同时,气候统计学在解决气候问题过程中亦有了令人惊喜的发展,新方法、新技术不断涌现,为气候研究注入了生机与活力,并已渗透到气候监测、诊断及预测的各个方面。应该说,应用统计方法研究气候问题的范围比以前拓宽了许多,除气候变化规律及预测的一般性研究外,区域性和全球性气候变化归因问题、气候数值模拟的检验、气候可预报性、运用不同时刻多个初始场的集合预报以及多个独立预报的集成等问题都运用了现代统计技术,且已取得了令人瞩目的成果。

我国的气候统计始自本世纪 60 年代。自那时以来,我国一直站在世界气候学术研究的前沿,至今已出版了 10 余本论著,其数量和学术内容均在国际学术界的前列。但近年来,由于计算机在气候统计研究方面应用的发展以及缺少对气候统计理论研究的投入,在气候统计研究和应用方面似有落伍之势态。事实上,由于气候的固有特性,不管使用什么方法,即便气候动力学和数值模拟,也需用现代统计技术加以处理。

魏凤英编著的《现代气候统计诊断与预测技术》一书较系统地介绍了气候统计诊断和预测方法,内容新颖,实用性强,更有理论,其中多维最大熵谱、Lepage 突变检测、主振荡型分析等内容是其它同类书籍中未曾涉及到的,尤其可贵的是,书中还论述了作者本人发展的若干新方法。此书以应用为目的,

以计算机计算为依托,详尽介绍了每种方法的原理及计算,并给出了作者从事气候研究中积累的大量应用实例,可以帮助读者更好地应用这些方法去解决气候研究的实际问题。我衷心祝贺本书的出版,以飨读者,更为我国现代气候统计的发展奠定基础。

么枕生

1999年6月

前 言

气候变化和异常已成为当今科学研究的重大课题之一,受到国家前所未有的重视,乃至成为公众街头巷尾、茶余饭后议论的话题。使用气候动力理论对气候变化和异常进行模拟、诊断,藉此达到预测气候的目的,构成了气候研究的主流方向之一。与此同时,人们惊喜地发现,气候统计诊断与预测技术亦有了长足的发展,近两年更呈上升趋势。正在不断推出的富有成效的新观点、新方法给统计诊断和预测带来新的生机,在气候研究和气候预测业务中起到举足轻重的作用。广大气候科研和业务人员迫切希望掌握这些新技术。为此,在 1997, 1998 和 1999 年,由中国气象科学研究所和北京气象学院联合举办了三期全国性“气候统计诊断与预测”讲习班,收到十分满意的效果。许多省、市也相继开办了类似的培训班。遗憾的是,至今还没有一套较为完整的教材,这给进一步理解和应用这些方法带来一定困难。为此,我萌发了撰写本书的想法。现在这一想法终于得以实现,将本书敬献给关心气候的学者、同仁和各界人士。

众所周知,在计算机进入概率论和数理统计领域后,概率统计出现了重大飞跃,出现了统计计算的新分支,它基于统计数学理论,并为解决在各个学科、各个领域提出的统计问题而不断发展。我们将此称为现代统计。本书的宗旨就是把近 20 年来在气象领域发展的现代统计技术介绍给读者。除介绍国外发展的新统计方法外,还着重介绍了作者自己发展的若干气候预测新方法。一部分必须介绍的常用经典方法,也尽量注入新的分析思想,从新的角度加以认识和使用。为节省篇幅,本书着眼于实际应用,没有详尽地给出这些方法的数学原理,但列出了使用所必须的数学公式。从解决气候研究和应用

中的实际问题出发,以计算机计算为依托,单刀直入地介绍每种方法的基本思想、原理、适用对象、计算步骤及计算结果分析要点。书中给出的大量应用实例,均是作者从事气候统计诊断与预测研究中长期积累的或是为编著本书专门计算的,力图通过这些算例为读者理解和使用这些方法提供帮助。在不少算例中给出了原始数据,便于使用者在计算机上实践操作。尽管本书属大气科学范畴,但书中介绍的诊断、预测技术可以很容易地“移植”到海洋、地震、环境、水文及生态等学科领域。

本书首先对气候统计诊断与预测的基本概念、包含的内容、基本统计量及统计检验等进行必要的介绍。然后依气候统计诊断与预测性质分类,对方法逐一进行叙述。

承蒙中国气象科学研究院曹鸿兴先生审阅了全书的初稿。在写作过程中,曹鸿兴和张先恭先生自始至终给予我热情的鼓励、悉心的指导和积极的帮助,对此我表示深深的谢意。我还要感谢北京气象学院江剑民、王永中、郑祖光先生及我的同事朱文妹、张光智、谷湘潜等,他们为本书的完成曾提供过种种帮助或有益的建议。衷心感谢我国著名气候学家、气候统计学界老前辈么枕生教授为本书作序。

由于作者水平有限,书中错误在所难免,恳请读者指正。

21 世纪的大门正向我们敞开。21 世纪必将是一个以科技素质为主导的竞争社会,只有拥有先进的科学技术,才会占领社会的制高点。正如变幻纷繁的世界一样,地球上的气候在 21 世纪也将是多变的,气候的变化也将深深影响着人类社会。对气候的监测、诊断和预测也会永不停息。愿本书能为研究 21 世纪的气候作出一点微薄的贡献。

魏凤英

1999 年 6 月

目 录

序

前言

第一章 绪论	(1)
§ 1.1 气候统计诊断概述	(1)
§ 1.2 气候统计预测概述	(6)
第二章 基本气候状态的统计量	(15)
§ 2.1 中心趋势统计量	(15)
§ 2.2 变化幅度统计量	(17)
§ 2.3 分布特征统计量	(19)
§ 2.4 相关统计量	(20)
第三章 基本气候状态的统计检验	(27)
§ 3.1 统计检验概述	(27)
§ 3.2 气候稳定性检验	(29)
§ 3.3 相关性检验	(35)
§ 3.4 分布的统计检验	(38)
第四章 气候变化趋势分析	(42)
§ 4.1 线性倾向估计	(43)
§ 4.2 滑动平均	(47)
§ 4.3 累积距平	(49)
§ 4.4 五、七和九点二次平滑	(50)
§ 4.5 五点三次平滑	(53)
§ 4.6 三次样条函数	(54)
§ 4.7 变化趋势的显著性检验	(59)

第五章 气候突变检测	(62)
§ 5.1 滑动 t -检验	(63)
§ 5.2 Cramer's 法	(65)
§ 5.3 Yamamoto 法	(66)
§ 5.4 Mann-Kendall 法	(69)
§ 5.5 Pettitt 方法	(72)
§ 5.6 Lepage 法	(73)
第六章 气候序列周期提取方法	(77)
§ 6.1 功率谱	(77)
§ 6.2 最大熵谱	(82)
§ 6.3 交叉谱	(88)
§ 6.4 多维最大熵谱	(95)
§ 6.5 奇异谱分析	(100)
§ 6.6 小波分析	(106)
第七章 气候变量场时空结构的分离	(114)
§ 7.1 经验正交函数分解	(115)
§ 7.2 扩展经验正交函数分解	(122)
§ 7.3 旋转经验正交函数分解	(128)
§ 7.4 复经验正交函数分解	(134)
§ 7.5 主振荡型分析	(141)
§ 7.6 循环稳态主振荡型分析	(148)
§ 7.7 复主振荡型分析	(151)
第八章 两气候变量场相关模态的分离	(155)
§ 8.1 奇异值分解式定理及其计算	(156)
§ 8.2 典型相关分析	(158)
§ 8.3 BP 典型相关分析	(169)
§ 8.4 奇异值分解	(173)

§ 8.5 SVD 与 CCA 及有关问题的讨论	(182)
第九章 最优回归预测模型	(187)
§ 9.1 多元线性回归的基本方法	(188)
§ 9.2 最优子集回归	(194)
§ 9.3 主成分回归	(201)
§ 9.4 特征根回归	(205)
§ 9.5 岭回归	(209)
第十章 均生函数预测模型	(214)
§ 10.1 均值生成函数.....	(214)
§ 10.2 双评分准则.....	(218)
§ 10.3 均生函数预测模型.....	(223)
§ 10.4 模糊均生函数模型.....	(229)
§ 10.5 全国夏季降水趋势分布预报方法.....	(232)
§ 10.6 最优气候均态模型.....	(240)
附录 1:现代气候统计诊断与预测程序使用说明	(245)
附录 2:附表	(260)

第一章 绪论

本章对气候统计诊断和预测的基本问题、内容和方法分别作一概述,对它们之间的相互联系作一鸟瞰式介绍。这些内容是阅读本书的基础。

§ 1.1 气候统计诊断概述

1.1.1 气候统计诊断的含义

这里给出有关名词的含义。

诊断 “诊断”一词源于医学。医生通过对病人的了解和检查。比如:中医用询问、切脉、看舌苔等办法;西医则用测压仪、X 光透视、超声波等仪器检查,从而判断病人所患何种疾病及所患疾病的原因、部位、性质及其病情程度,这一过程称为诊断。

统计诊断 “统计诊断”是指对统计建模及统计推断过程进行诊断。它是 20 世纪 70 年代中期才发展起来的一门统计学新分支^[1]。统计诊断是对收集起来的数据、以数据为基础建立的模型及相应的推断方法的合理性进行分析。通过一些统计量来检查数据、模型及推断方法中可能存在的“病患”,提出“治疗”办法。为了克服模型与客观实际之间可能存在的差异,需要寻找一种诊断方法,判断实际数据与模型之间是否存在较大偏离,并采取相应对策。这就是数理统计意义下诊断的基本内容。通过统计诊断,找出严重偏离模型的异常点,区分出对于统计推断影响特别大的强影响点。其中对多元线性回归

的诊断是统计诊断的主要内容之一。

气候统计诊断 将“诊断”一词引入到气候学研究中,用某些手段根据气候观测资料对气候变化与气候异常的程度与成因作出判断,即称之为气候诊断。由于是用统计手段进行气候诊断,故将这种诊断称为气候统计诊断。

可见,气候统计诊断的含义与统计学分支——统计诊断的含义是不同的。前者是用统计学方法对气候过程进行诊断,而后者是对统计建模与统计推断过程中可能出现的问题进行诊断。

另一方面,气候统计诊断与通常的气候统计分析也有一些区别。气候统计分析是根据大量气候资料用概率论与数理统计中的方法,研究气候演变的时空变化特征和规律。气候统计诊断除了进行以上一系列分析外,还要进行一系列的科学综合和推断,期望通过统计方法这一气候诊断的重要手段对气候变化与气候异常及其成因作出正确判断^[2~3]。例如,气候研究中所谓气候变化归因问题,就是气候诊断的重要内容。在实际应用时,气候统计分析与诊断又往往很难区分开。

1.1.2 气候统计诊断研究的内容

概括地讲,采用统计方法进行气候诊断研究,主要包括以下几方面内容:

(1)应用统计方法了解区域性或全球性气候变化的时空分布特征、变化规律及气候异常的程度。主要针对月以上至几十年时间尺度的变化,即主要研究月、季、年及年代4个时间尺度的气候变化。

(2)通过统计方法探索气候变量之间及与其它物理因素之间的联系,以此研究造成气候异常的原因,进而探索气候异常形成的物理机制。

(3)对气候数值模拟结果与实际变化状况之间的差异进行统计诊断。

1.1.3 现代气候统计诊断技术窥视

气候诊断使用的统计技术涉及到统计学多个分支,如统计检验、时间序列分析、谱分析、多元分析、变量场展开等等。近年来,在引入气候模式对气候异常和变化进行诊断的同时,气候统计诊断技术也有了长足的发展,引起气候工作者的关注。作者之所以在书名“气候统计诊断”前面加上“现代”一词,用意在于希望本书能够尽量反映国际气候统计诊断技术的现代水平。如同医学诊断一样,随着科学技术的发展,现代诊断技术与经典诊断技术有了显著的不同。新概念、新方法不断涌现,逐步取代了原有的经典诊断方法,为气候诊断研究提供了更科学、更有效的手段,同时也拓宽了人们认识气候系统的视野^[4~6]。与经典方法相比,现代统计诊断技术的发展主要体现在如下几方面:

1.1.3.1 气候变化趋势和突变检测

从气候序列中分离气候变化趋势,不仅采用滑动平均、累积距平、线性倾向估计等传统方法,还引入了样条函数等数据拟合的新方法。采用这些方法对气候序列的分段曲线拟合,以便更好地反映其真实的变化趋势。此外,还注重对变化趋势进行显著性检验。

尽管目前突变统计诊断技术还很不成熟,但是针对突变问题,借助统计检验、概率论等发展了一些行之有效的检测方法。例如,气候均值、变率以及事件发生与否的检验。气候诊断研究中使用最多的是均值的统计检验。其中不但使用参数统计检验,而且还使用非参数统计检验。

1.1.3.2 气候振荡分析

近年来,诊断气候振荡的技术发展很快。从不连续的周期图、方差分析、谐波分解发展到连续谱、一维、多维最大熵谱。近年来,又发展了动力重构与经验正交函数相联系的新技术——奇异谱,以及能将不同波长的波幅一目了然地展现在一张二维图像上的小波变换。这些新技术与传统技术相比,分辨率更高、适用性更强,对于揭示气候序列不同时间尺度的振荡特性起到了很大的作用。

1.1.3.3 气候变化时空结构诊断

以经验正交函数为基础的对气候场进行时空分布特征的诊断技术也有了令人瞩目的发展。针对气候变量场特征分析的不同需要,发展了揭示变量场移动性分布结构的扩展经验正交函数、着重表现空间的相关性分布结构的旋转经验正交函数、可以展示空间行波结构的复经验正交函数和描述动力系统非线性变化特征的主振荡型、循环平稳(Cyclostationary)主振荡型、复主振荡型等等。

1.1.3.4 气候变量场间耦合特征诊断

气候研究中常常遇到两个变量场的相关问题。例如,相隔遥远的不同区域同一时间或不同时间变量场间存在的遥相关、海洋与大气的相互作用、大气环流或下垫面对气温和降水的影响等等问题。过去讨论这些问题多用相关分析。现在将典型相关这一有着坚实的数学基础、推理严谨的两组变量分析工具移植到两变量场耦合特征的诊断中。还提出了从讨论两个场主分量出发的 BP 典型相关分析。同时,奇异值分解也在两场耦合特征研究中广泛使用。

1.1.4 气候统计诊断的一般步骤

利用统计方法进行气候诊断,一般可分为下列几个步骤:

1.1.4.1 收集资料

从研究的实际问题出发,确定统计诊断的对象,收集有关的资料。选取的资料应该准确、精确,并具均一性、代表性和比较性。资料的样本长度和区域范围与所研究问题的对象有关。例如,研究中国气候年代际变化规律,应该收集半个世纪至百年以上的样本;研究准两年振荡则有数十年样本就够了。在研究气候场空间结构时,选取某一区域范围内的站点布局应该满足均一性和比较性,否则不能很好地反映变化的真实状况或诊断结果缺乏代表性,且难以比较各区域各时期气候特征的差异。

1.1.4.2 资料预处理

对于收集起来的资料,根据研究问题的具体需要进行预处理。各个气候变量的单位不相同,平均值和标准差亦不相同。为了使它们变为同一水平无量纲的变量,通常要对资料进行标准化处理。标准化的变量均值为0,方差为1。有时根据实际情况对资料作距平化处理,给研究带来便利。

1.1.4.3 选择诊断方法

根据研究目的和研究对象,选择合适的诊断方法进行研究。选择不恰当的方法会给研究和物理解释带来困难。例如:研究大气准两年振荡的时空演变特征,将变量场进行带通滤波后,使用扩展的经验正交函数可以展现出不同位相准两年振荡的变化。对于这种研究目的,其它方法则无法做到。再例如:划分气候区域的研究,使用旋转经验正交函数,按照分离出旋转典型空间模态的高荷载区可以进行客观的区域划分,而使用普通的经验正交函数很难收到理想的效果。

1.1.4.4 科学综合和诊断

气候统计诊断是统计学与气候学间的交叉学科。利用统

计方法进行气候诊断,不能陷入盲目套用计算公式。在一些情况下,对计算结果应该进行显著性检验。没有统计意义的结果是失真的,没有分析的价值。这一点往往被人们所忽视。要得出科学的结论,重要的是运用深厚的气候学知识,对计算结果进行科学的综合和细致的分析。如同诊断疾病,统计计算结果好比 X 光透视片子或超声波图像,要确定所患何种疾病及其部位、性质、病情程度,需要医生凭借医学知识和临床经验,对这些检查结果进行综合分析,才能得出正确的诊断,这是医术是否高明的重要标志。同样在气候统计诊断中,对统计计算结果需用气候专业知识进行判断、识别真伪,概括出气候系统确实存在的事实以及彼此间的联系。

§ 1.2 气候统计预测概述

1.2.1 气候统计预测的一般概念

按照统计学的观点,利用统计模型估计随机动态系统未来可能出现的行为或状态,称为统计预测。具体地讲,统计预测是利用历史与现时的观测值 $x_1, x_2, \dots, x_{t-1}, x_t$ (t 为现在时刻),估计这个随机系统未来 m 个时刻的状态值 $x_{t+1}, x_{t+2}, x_{t+m}$ 。可见,统计预测是以系统的过去和现在的信息为基础,对未来时刻作出估计。利用统计模型对气候系统的未来变化状态作出估计,即为气候统计预测。当然,统计模型是在利用大量过去气候资料对气候系统内部或与其它变量之间关系的变化规律及特征的分析基础之上建立的。

1.2.2 气候统计预测的基本假设

在使用统计模型对气候系统未来状态进行统计预测时,隐含着一个基本假设——气候系统的未来状态类似于过去和

现在。这一假设体现在利用统计模型对未来状态进行预测时，是假设模型结构在预测期间内保持不变、气候系统变化及与各变量之间的相关关系在预测期间不变。

从统计学理论上讲，气候统计预测的基本假设应该满足以下两个条件：

(1)气候变化的成因和物理机制至少在预测期间与观测时期一致。

(2)气候系统在预测期间保持稳定。由于气候系统具有一定的概率特性、因果特性和相关特性，因此气候预测很大程度上依赖于统计预测。但是，由于统计预测是在假设系统未来仍按过去和现在的特性变化的前提之下，一旦气候系统出现异常甚至突变或影响气候系统的因素有所改变时，往往导致预测失败。因此，较高技巧的预测源于对气候系统变化特性的深刻了解和认识。

1.2.3 气候统计预测的基本要素

气候统计预测过程主要由以下几个要素构成：

1.2.3.1 预测对象

预测对象是指欲预测的气候要素。比如对某区域旱涝趋势、冷暖趋势或夏季降水量或某月气温等等进行预测。可以是某一测站的局地预测，也可以是大范围区域乃至全球性预测。

1.2.3.2 预测依据

在气候系统内部或影响其变化要素相互关系的诊断基础上提供的预测依据。通常为从某些统计上显著相关的预报因子群中提取的有效信息。

1.2.3.3 预测技术

根据数据性质及预测对象、预测因子的特点，选择合适的统计预测模型。

1.2.3.4 预测结果

对未来气候变化状态时间、空间、数量、性质等方面的预测。

在以上四个基本要素中,第三个要素包含的内容最丰富。

1.2.4 预测技术窥视

统计预测技术形式繁多、分类方式也是多种多样的。若按照预测性质划分,预测技术可以分为两大类,即定性预测与定量预测。

1.2.4.1 定性预测

定性预测方法主要依赖气候专家的主观认识能力,综合地分析过去、现在和将来可能出现的各种因素之间的相互影响,寻找气候要素的发展规律,对未来的发展趋势和性质作出推断。例如:气候学家根据气候学知识,对海温与副热带高压、青藏高原的热力作用、西风带环流及东亚季风等因素的过去、现在和未来可能出现的状态进行综合分析,寻找它们影响中国夏季气候异常特别是降水异常的演变规律,对未来夏季气候是否异常、异常的程度如何作出定性的推断。

定性预测纯数学的处理手段较少,所需资料数量也不必很多,但并不意味着定性预测不需要数量分析。除气候学专业知识和预测经验外,仍然需要一定的数量分析和统计处理,使预测更科学、更可靠。例如:气候预测专家们在对1998年全国夏季降水趋势作预测时,抓住了几个影响我国气候的主要因素:1997~1998年间热带太平洋海水出现异常增温、青藏高原出现空前的大雪、东亚冬季风异常偏弱、西太平洋副热带高压异常偏强等等。根据对这些异常现象的分析,得出1998年夏季长江中下游及江南北部可能出现严重洪涝的预测。其实,这个定性预测中包含了一定的统计处理。海温出现异常,确定