

关于统计气象学的几个问题

周 家 斌

(中国科学院大气物理研究所, 北京 100029)

提 要

本文讨论了当代统计气象学的发展, 提出了作者认为值得重视的几个问题。

关键词: 统计气象学。

一、引 言

第五届国际统计气候学会议 (5th International Meeting on Statistical Climatology) 于 1992 年 6 月 22 日至 26 日在加拿大多伦多召开。美国第十二届概率统计会议 (12th Conference on Probability and Statistics in the Atmospheric Sciences) 与该会议联席举行。出席会议的有一百多人。会议收到论文 184 篇, 其中 26 篇为联席会议论文。我国学者向会议提交了 19 篇论文, 5 人参加了会议并作了学术报告。

作者从有关方面获悉, 一个关于统计气候学的国际性协会即将成立。这对我国统计气象学界来说, 是机遇也是挑战。我国统计气象学有相当高的水平。我们应当更多地参加国际学术交流, 争取更快地发展。

本文的目的在于结合多伦多会议, 谈一谈作者认为现代统计气象学发展中值得重视的几个问题。

二、统计学方法在业务预报中的作用

目前短期天气形势预报基本上采用数值预报方法, 而气象要素预报则更多地依赖统计学方法。一个行之有效的方法是应用数值预报产品的模式输出统计 (Model Output Statistics, MOS) 方法。MOS 方法尚存在一些问题, 主要是该方法的预报模型和数值模式的性能密切相关, 因此当数值模式改变时预报模型也需要改变。近年来数值模式更新换代速度加快, 使这一矛盾更为突出。为了使预报模型适应数值模式的变化, 近年来有人将 Kalman 滤波技术引入 MOS 预报之中^[1]。

一般的多站预报都是利用逐站建立预报方程的方法制作的, 没有考虑面上要素分布的整体性质。如何直接做要素分布预报, 是一个值得研究的问题。各种降维技术在这方面大有用武之地。

目前,中期数值预报方法所做的形势预报,一周内有应用价值,再长则误差较大,需用统计方法补充订正。至于要素预报,则仍以 MOS 方法为主。

长期数值预报近年来有了很大发展,但目前还难以建立业务预报,实际业务预报仍以统计方法为主。经过多年的预报实践,国内外都已逐渐认识到充分反映长期天气过程物理规律和预报员经验的物理统计方法在长期天气预报中的重要地位。世界气象组织经过对许多地方长期天气预报效果的统计,确认了物理统计方法在长期天气预报中的可行性。因此,如何将动力学和统计学相结合,并在此基础上建立新的预报模式,是很值得探讨的。非线性动力学在这方面可以发挥重要作用。

三、天气气候学统计诊断分析

天气气候学分析是统计学方法在气象学中应用的一个重要方面。以往的天气气候学分析,多限于个例分析。应用现代统计学方法对长期资料的统计诊断,大大提高了天气气候学成果的可靠性和普遍性。

下面简略叙述一下气候统计诊断的几个方面。

首先是气候资料分析。以温室效应为例,我们首先遇到的是全球平均温度的计算。全球平均温度的计算结果受到测站分布不均匀、资料误差(器测误差、地形影响等)、资料长度、观测仪器及时次变化的影响,因而需要作为专门的统计学问题进行讨论^[1]。

求出全球平均温度的时间序列之后,尚需对温度是否有上升趋势和这种上升趋势与二氧化碳排放量的关系进行统计检验,才能对温室效应问题给出一定的回答。

气候突变是近年来十分引人注目的研究课题,从非线性动力学的观点看,突变是系统从一个平衡态向另一个平衡态的跃变。从资料看,表现为突变前后平均值的改变。另外,还有其他统计性质(如方差)的改变等突变方式。识别突变的方法很多。近年来应用得较多的是滑动 t 检验法和 Man-Kendall 方法^[2],另外,还有 Petetitt 方法^[3]。这两种方法的统计意义是突变点以后的时间序列稳定地大于(或小于)突变点以前的值。杨喜寿等^[4]给出了一种检测气候突变的逐步算法。

这些算法主要是检测均值突变。至于方差突变等则需另作研究。这些方法着眼的是单变量时间序列,如何从多要素综合特征的变化上研究气候突变则值得探讨。除了要素时间变化特征的突变外,尚有气候系统空间结构特征(如旱涝分布)的突变。另外,超熵的产生也可以作为气候突变的一个判据。

气候可预报性对于了解气候变化规律、研究预报方法有重要意义。短期天气的可预报性一般指初始误差的增长所导致的预报误差的增长。数值预报的实践指出,可用预报的时限大约为两周。实际上,上述误差不能认为是完全由初值误差增长造成的,它与模式性能(包括动力框架和计算方法)等很有关系。亦即我们应该区别大气的可预报性和模式的可预报性。大气可预报性指初始误差由于大气过程的非线性性质而逐步增长所导致的可预报时间尺度,模式可预报性指模式性能引起的误差增长而产生的可预报时间尺度。

1) 杨喜寿、杨洪昌,气候变点的推断。

前者可通过资料的统计分析研究,后者则通过模式产品的统计分析研究。当然要完全地把大气的可预报性从模式产品中分离出来是相当困难的。

大气是一个非线性系统,它对初值的变化十分敏感。描述这种非线性系统演变特征的一个指标是系统的 Lyapunov 指数。应用实际资料可以计算各种尺度天气系统的 Lyapunov 指数,从而得到它们的可预报时间尺度^[5]。

一般认为,短期天气变化是大气内部动力过程引起的,而长期天气变化则是由大气外部的动力与热力影响通过大气内部的变化产生的。大气过程引起的短期天气变化则可看成叠加于长期天气过程上的噪声。关于长期天气可预报性的研究,主要就是分析大气外部缓慢变化产生的长期变化(即气候信号)与大气动力过程引起的短期变化(即气候噪音)之间的关系。人们常用信号与噪音之比(简称信噪比)来度量可预报性。

气候系统有不同的时间空间尺度,不同尺度的运动可以看成不同的层次。例如,月、季尺度、年际尺度、年代际以上尺度可以看成不同的气候层次。高层次的运动可处理成低层次运动的背景场,低层次的运动则作为对高层次的扰动。通过资料分析,可以建立有关气候层次变化的方程,其中低层次运动的影响常用随机扰动表示^[6]。

四、数值模式的统计研究

数值模式在大气科学中起着重要作用。其主要应用有气候模拟、气候敏感性试验、天气预报等。数值模式的基本框架是动力学方程。但在数值模式的设计、运行和结果的评价和应用中无处不用到统计学方法。

数值模式在设计过程中需要解决物理因子选择、描述、物理参数确定等问题。在模式中需要考虑哪些因子(或方程中的哪些项),应当立足于实际资料的统计分析。关于某些物理过程的描述方法,有关物理参数的确定,以及某些物理过程的参数化方案,亦需根植于资料的统计分析。

数值天气预报系统的各个环节,如资料质量控制、客观分析、初值形成及资料同化、数值计算、结果输出及应用,无不渗透着多种统计方法如插值、优化、决策论等。

目前长期数值天气预报中所用的综合平均预报(Ensemble Average Forecast)和滞后平均预报(Lagged Average Forecast)实际上是一种简单的算术平均方法,有待于进一步的发展。

关于数值模式结果的评估,包括模拟结果与实际观测的比较、敏感性试验结果分析、预报评分等方面都需要先进的统计学方法。例如,模式结果与实际观测的比较、模式之间的相互比较,需要借助于统计检验。个别或少数例子的直观比较其可靠性较差,只有相当数量结果的统计检验才有说服力^[7]。国内目前关于数值模式结果的讨论由于种种原因限于个别或少数个例,极需开展大量个例的统计评估。数值模式的结果是一个场,因此它的评估是一个多变量统计问题。在这方面统计学方法本身尚不够完善。某些新发展起来的非参数方法——如多因子排列法(Multi-Response Permutation Procedures)值得试验。

气候模拟关心的是平均状态,而将方差看成噪音。气候变率就是一种方差,它在气候学研究中有重要地位。关于数值模式的统计评价常常限于平均值的统计检验,而对方

差的比较重视不够。在这方面需要作进一步的研究。

关于数值模式的统计研究,目前在国内仍是一个薄弱环节,亟待加强,以促进数值模式研究的发展。

五、降维技术

气象资料的时空分布非常复杂,如果直接分析原始资料,难以找出其中的规律,其原因之一是维数太高。因此如何将资料压缩,分析其基本特征,就是降维技术所要解决的问题。

谱分析是一种传统的降维技术。它在揭示大气中存在的各种周期性波动中起着重要作用。利用非整数波分析技术,对于揭示长周期的低频波有相当好处。奇异谱技术的应用,有利于分析优势周期。各种截谱模式的应用,对大气过程物理机制的研究和数值天气预报起着重要的作用。然而,谱分析基底函数所表示的基本波型,在大气中比较少见,常常难以给出分析对象的基本特征。

车贝雪夫展开在表示区域大气运动特征方面有其优越性。但是,传统的车贝雪夫展开只能用于矩形等距网格。作者将车贝雪夫展开推广到不规则格点上^[8]。推广后的车贝雪夫展开方法既能用于不规则格点而其基底函数又不随时间变化,因而在气象要素场的分析与预报方面有着更加广泛的用途。作者^[9]提出的基于车贝雪夫展开的迭代算法,是一种新的时间序列预报方法。它不含平稳性假定,因而是一种非参数方法。这个方法近年来得到了进一步的推广^[10]、改进和简化^[11]。

经验正交展开方法可以处理不规则格点的资料,且收敛较快,因而经常用于气象场分析。由于它的基底函数随时间变化,因而不宜用于预报。近年来的研究发现,它对于揭示不同地域变化特征方面有不足之处。因而又发展出旋转的经验正交展开。复经验正交展开可以给出场的空间振幅和位相以及时间振幅和位相,因而便于分析场的时间变化^[12]。

主振荡模态 (Principal Oscillation Patterns) 方法将多维时间序列看成马尔可夫过程,以其时滞相关系数矩阵的特征向量进行正交展开,所获得的基底函数代表性较好,可以分析 30 至 50 天低频振荡,也可以分析大气运动的不稳定性,还可以用来做预报。近年来有了复主振荡模态分析^[13],这一技术在国内还很少研究,值得引起重视。

投影追踪方法 (PP) 也是一种降维技术,它具有稳健性的优点。这个方法现已发展出 PP 主成份分析、PP 判别、PP 聚类、PP 回归、PP 多元参数估计、PP 时间序列分析、PP 密度估计等。

反映气象要素间非线性关系的尺度分析 (Scaling) 是降维技术的新发展,它还具有可以更好地吸收诊断分析成果的优点。这个方法在理论上还不够完善,在我国研究也不多,值得做进一步的研究。

各种降维技术有某些共同特征,也各有自身的特点,有些技术之间 (如主振荡模态方法和谱分析) 还有一定的联系,很值得做一些综合性研究。

六、非线性动力学

自从 Lorenz 关于确定性非周期流的论文发表以来, 非线性动力学的理论和应用有了很大发展。除了从简化的低阶谱模式出发对大气过程作理论性分析研究之外, 关于用非线性动力学观点处理历史资料的研究也有了广泛开展。这方面的应用研究主要有两个方面。其一是计算系统的维数。若维数是分数, 则认为系统具有混沌性质。已有分析结果表明, 从小尺度的湍流运动到 ENSO 这样的短期气候过程, 其维数大都在 5 至 8 之间, 这说明大气运动是一个具有有限个自由度的确定性的混沌系统。另一是计算最大的 Lyapunov 指数, 确定最大的可预报时间尺度; 计算二阶熵, 确定平均可预报时间尺度。上述问题的计算方法尚不够多, 有待于进一步发展。已有的研究成果如何进一步深入也值得探讨。目前只能处理单变量的时间序列, 所用信息不够丰富。如果能发展出一套多维时间序列的计算方法, 将可使这一领域的研究大大前进一步。

作者认为, 如何将混沌概念用于预报, 也应当是应用研究的一个重要方面。可惜目前这方面的研究不多。1987 年 Farmer 等^[13]发表关于预报方法的第一篇论文以来, 国外这方面的论文甚少, 讨论也不够深入。林振山^[14]首先在国内开展了这方面的研究, 提出了几个相空间预报模式。作者近年来也就此进行了研究, 提出了四种以混沌理论为基础的预报方法, 即相空间向量相似方法、Lyapunov 指数方法、空间变换方法和模序列预报方法^[15]。其中一部分已用于南方涛动强度、北京降水和华北降水分布的预报。

混沌概念的提出, 打破了确定论和随机论的鸿沟, 给许多传统的科学概念以很大冲击。经典的动力气象学和统计气象学都将在某种意义上被改造。混沌概念进入统计气象学, 必将促进这一学科的发展。

七、结 束 语

本文讨论了当代统计气象学的若干问题, 这些问题是作者认为值得进一步关注的。这当然不能概括统计气象学的所有领域。因而本文不能看作是统计气象学的全面评述。

今后我们仍然需要把在业务预报中应用统计学方法放在重要地位。一个方法只有在业务预报中扎下了根, 才能具有永久的生命力。

统计气象学的研究领域需要不断拓宽, 在发展数值预报和大气环流模式中充分运用统计学的成果, 将会使这些领域获得更健康的发展。在大气物理学领域, 统计学将是大有作为的。

需要不断引进和研究新的统计学方法。有些方法, 如主振荡模态分析、尺度分析、神经网络方法、各种非参数方法、非线性动力统计模式等, 有的尚未引入我国, 有的研究尚不充分, 需要进一步发展。

结合不同的领域发展统计气象学, 仍是今后相当一个时期的重要方向。应用各种先进的统计学方法, 将会使大气科学各个分支的研究不断深化, 特别是天气学和气候学及气候预测, 仍将是统计学大有作为的领域。

如何进一步发展统计气象学,是一个值得探讨的问题。统计气象学需要有自己的基本概念、独特方法、基本内容和完整体系。今天的统计气象学,离这个目标还相当远。但是,经过持久的努力,这一目标是一定能够实现的。

参 考 文 献

- [1] Persson A. O., 1991, Kalman Filtering-A New Approach to Adaptive Statistical Interpretation of Numerical Meteorological Forecast, WMO/TD, No. 421.
- [2] Shen S. S. et al., 1992, Sampling Errors in Estimating the Global Temperature, Preprints, 5th Inter. Meeting Statis. Climat., Toronto, 419—420.
- [3] Goossens C. H. et al., 1986, Annual and Seasonal Climatic Variations over the Northern Hemisphere and Europe during the Last Century, *Ann. Geophys.*, **4**, 385—400.
- [4] Pettitt A. N., 1977, A Non-parametric Approach to the Change-point Problem, *Appl. Statist.*, **28**, 126—135.
- [5] 杨培才等, 1990, 埃尔尼诺/南方涛动的可预报性, *大气科学*, **14**, 64—71.
- [6] Liu Shida, Lin Zhenshan, 1989, Application of Climatic Hierarchy Theory for the Study of Modern Climate, *Ann. Rep.*, LASG, 467—477.
- [7] Kaltz, K. W., The Role of Statistics in the Validation of General Circulation Models, *Climate Res.*, (in Press)
- [8] 周家斌, 1990, 车贝雪夫多项式及其在气象中的应用, *气象出版社*.
- [9] 周家斌, 1991, 时间序列预报方法的推广, *大气科学* **15**, 73—81.
- [10] 周家斌、朱文琳, 1992, 时间序列预报方法的进一步简化, 长期天气预报和日地关系研究, *海洋出版社*, 157—164.
- [11] 黄嘉佑, 1988, 准两年周期在我国降水场中的表现, *大气科学*, **12**, 267—273.
- [12] Burger G., Complex Principal Oscillation Pattern Analysis, Rep. MPI, No. 86.
- [13] Farmer J. D. et al., 1987, Predicting Chaotic Time Series, *Phys. Rev. Lett.*, **59**, 845—848.
- [14] 林振山, 1993, 长期预报的相空间理论和模式, *气象出版社*.
- [15] 周家斌、王允宽、杨桂英, 1993, 四种以混沌理论为基础的预报方法, *数理统计与管理*, 增刊, 229—233.

Some Problems in Statistical Meteorology

Zhou Jiabin

(Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029)

Abstract

Current status of statistical meteorology is discussed. Some problems which needs to be emphasized are raised.

Key word: Statistical meteorology.