

朱玉祥, 黄嘉佑, 丁一汇. 2016. 统计方法在数值模式中应用的若干新进展. 气象, 42(4): 456-465.

统计方法在数值模式中应用的若干新进展^{* 1}

朱玉祥¹ 黄嘉佑² 丁一汇³

1 中国气象局气象干部培训学院, 北京 100081

2 北京大学物理学院大气与海洋科学系, 北京 100871

3 国家气候中心, 北京 100081

提 要: 目前, 数值模式的准确性不断提高, 天气预报和气候预测越来越依靠数值模式。数值模式不仅仅是求解描述大气运动和热力过程的方程组, 其中也大量使用了统计方法, 统计方法是数值模式的重要组成部分。本文对统计方法在数值模式中应用的最新进展进行了综述。首先, 分析了统计方法在数值模式的资料同化、集合预报、次网格尺度物理过程参数化方案、数值产品统计释用、延伸期预报、模式检验等方面的应用。然后阐述了贝叶斯统计在数值模式中应用的新进展。最后, 对统计方法在数值模式中应用的未来发展趋势进行了展望。

关键词: 数值模式, 统计释用, 统计检验, 贝叶斯统计, 数据挖掘

中图分类号: P413.456.468

文献标志码: A

doi: 10.7519/j.issn.1000-0526.2016.04.009

Some New Advances in Application of Statistical Methods to Numerical Models

ZHU Yuxiang¹ HUANG Jiayou² DING Yihui³

1 China Meteorological Administration Training Centre, Beijing 100081

2 Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing 100871

3 National Climate Centre, Beijing 100081

Abstract: The accuracy of numerical models has been improved continuously and weather forecast and climate prediction rely on numerical models more and more. Numerical models are the equations which describe the motion and thermodynamic processes of the atmosphere. Statistical methods are also widely used to numerical models, becoming important components of numerical models. This paper reviews some new advances in the application of statistical method to numerical models. Firstly, application of statistical methods to data assimilation, ensemble forecast, parameterization of physical processes, statistical interpretation, extended-range weather forecasts and model verification for numerical model are analyzed. Then, the latest application of Bayes statistics to numerical models is elaborated. Finally, several important research trends for future application of statistical methods to numerical models are put forward.

Key words: numerical model, statistical interpretation, statistical verification, Bayes statistics, data mining

* 国家重点基础研究发展计划(973 计划)(2012CB957804)、公益性行业(气象)科研专项(GYHY201406020)、国家自然科学基金项目(41305131 和 41505079)及重庆市博士后项目(渝 xm201103028)共同资助

2015 年 8 月 31 日收稿; 2015 年 12 月 3 日收修定稿

第一作者: 朱玉祥, 主要从事气象统计方法的和青藏高原积雪的研究. Email: zhuyx@cma.gov.cn

引言

统计方法在气象科研和业务的各个领域具有广泛应用(魏淑秋,1985;么枕生等,1990;马开玉等,1993;丁裕国等,1998;施能,2002;2009;黄嘉佑,1990;2004;吴洪宝等,2005;魏凤英,2007;黄嘉佑等,2015;von Storch et al,1999;Wilks,2011)。其实,统计思想在气象中的应用具有悠久的历史。中华民族五千年的历史文明中蕴藏着丰富的农谚,如“不得春风,难得秋雨”和“八月十五云遮月,正月十五雪打灯”,这些农谚实质上就是早期经验性的气象统计研究,反映了气象统计在古代的应用。近代气象统计学的发展是与计算工具和数理统计学的进步密切相关的。

自20世纪80年代以来,随着计算机技术的迅猛发展,现代统计学突破了仅对观测资料统计指标的简单计算和分析的局限,出现了重大飞跃。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视角上。过去有一些文章对气象统计方法的应用进展进行了综述。比如,周家斌等(1997)对中国统计气象学的进展进行了综述,综述内容包括统计方法在业务预报中的应用、天气学、大气运动的自忆性方程、非线性动力学、经验正交函数、车贝雪夫展开、灰色分析与模糊预测、判别分析、动态数据处理与多层递阶方法、子波变换、非线性回归分析等方法在气象学中的应用;谢炯光等(2003)主要从多元统计分析和时间序列分析上选材,介绍了近30年来气象统计预报在中国气象业务科研中的一些应用和发展;Zwiers等(2004)综述了统计方法在气候研究中的应用;魏凤英(2006)概述了近50年来中国气象科学研究院在气候统计诊断与预测方面的研究进展。这些综述气象统计方法进展的文章大都写于十多年前。而最近十几年来,在气象科研和业务取得巨大进步的同时,统计方法在气象科研与业务中也出现了很多新应用,特别是数值模式的进步导致气象统计应用出现了很多新情况。因此本文对统计方法在数值模式中应用的新进展进行综述,并且对其未来发展趋势进行展望。

1 数值预报的内涵

天气、气候现象是地球大气运动的结果,它们受

一定的物理、化学定律的支配,这些定律可以用一组微分方程来表示。从一定的初始状态出发,在一定的环境条件(即边界条件)下求出这一微分方程组的解,就可能对未来的天气或气候状况做出预报。由于这些方程的复杂性,必须借助于现代高性能计算机,使用数值方法才能求解,这就是数值预报(矫梅燕,2010)。

数值天气预报的成功被认为是20世纪最重大的科技和社会进步之一。目前,数值模式的准确性越来越高,天气预报和气候预测越来越依靠数值预报的结果(矫梅燕,2010)。对数值模式产品的理解和应用能力也已经成为衡量预报员水平的重要标准。但在看到数值模式巨大潜力的同时,也需要对数值模式存在的问题与困难有充分的认识。数值模式存在的问题与困难主要表现为:(1)数值模式所描述的大气物理过程是有限的。比如人类的各种活动就无法在数值模式中精确描述,数值模式中地形的描述也与实际地形有较大差别,雷电在很多数值模式中也没有考虑。积云、辐射、边界层等各种参数化方案,也都是对大气过程的近似描述。(2)初始场不可能绝对准确。观测资料总是存在误差,观测资料的密度也是有限的,从观测资料插值到模式初始场(客观分析或资料同化)的时候也存在误差。(3)计算过程中的舍入误差无法避免。目前,对于描述大气运动的微分方程组人类尚无法给出精确的解析解,只能利用数值计算方法将该微分方程组离散为计算机可以求解的差分方程组,求其近似的数值解。同时,计算机都存在存储字长的限制,每一个变量只能用该种字长的数值来表示,所以计算过程中的每一步都有舍入误差。

气象研究方法大致可以分为动力方法和统计方法两类。动力方法是确定论的,动力方法认为现在的天气和气候现象是过去发展的必然结果,其演变过程可以用流体动力学方程描述,对一组确定的初值,其动力系统未来的演变过程是确定的。统计方法是基于概率论的,统计方法研究大量气象现象之间的关联及其自身的演变规律,这种方法把气象变量看作为随机事件,不去关注它们精确的个别演变行为,而去研究它们大量现象的总体行为,由此发现其中的规律性。上面描述的数值模式的问题与困难表明,期望数值模式采用纯动力的方法,得到完全准确的确定预报是不现实的,必须要充分吸收和发挥统计方法的优越性,动力与统计相结合才能做好天

气预报和气候预测。

当今,数值模式并不是单纯的动力学方法,其中也大量应用了统计方法,下面将对统计方法在数值模式中的应用进行分析和总结。

2 统计方法在数值模式中的应用

在数值模式初始场的资料同化、集合预报、次网格尺度物理过程参数化方案、模式检验、数值产品的统计释用等方面,都大量使用了统计方法(Kalnay, 2001;蒲朝霞等,2005;Rasmus et al,2008)。

2.1 初始场的资料同化

数值预报是在给定初值后积分数值模式获得未来时刻的预报预测信息。资料同化,就是为数值模式提供与观测最接近,并且与数值模式相匹配的初值。这里的“与观测最接近”是统计意义上的。统计方法是资料同化方法的理论基础,而其中最主要的是统计估计理论,最常用的是最小方差估计、最大似然估计和贝叶斯理论(邹晓蕾,2009)。资料同化技术经过了多项式插值、逐步订正、最优插值等历史发展阶段,目前业务中普遍使用的方法是三维或四维变分同化,集合卡尔曼滤波方法也在进行业务尝试中(矫梅燕,2010)。这些技术都离不开统计,比如变分同化是统计和动力相结合(邹晓蕾,2009);而最优插值和卡尔曼滤波是纯统计方法(蒲朝霞等,2005;邹晓蕾,2009;朱国富,2015a;2015b)。

由于在很多偏远地区和大洋区域,常规观测资料的数量和质量都十分有限。在这些区域长时间增加常规观测需要花费大量的财力和物力,并不可行。虽然气象雷达和卫星等遥感资料的应用可在一定程度上改善这些地区资料匮乏的状况,但这些遥感资料的质量尚不足以对高影响天气做出实时、精确的预报。针对这一问题,科学家们提出了一种被称为目标观测的方法,即在目标观测时刻将有限的观测资源增加至敏感区(穆穆等,2011)。敏感区是指初始误差的增长会导致预报准确度下降最快的区域。如何确定敏感区是目标观测的关键,气象学家提出了多种确定敏感区的方法,如集合卡尔曼滤波、奇异向量、条件非线性最优扰动等(Mu et al,2009),这些方法大都是统计方法。

综上所述,数值模式初始场的形成离不开统计方法。

2.2 集合预报

由于数值预报所需要的大气初始状态只能近似的确定,因而对天气预报问题的完全描述,应该提为大气运动相空间中大气状态的概率密度函数(PDF) $\rho(X, t)$ 的预报,其中 X 为气象变量, t 为时间。该问题在理论上可表述为概率的连续方程,即 Liouville 方程(Palmer et al,2006)。但是,即使在只有几个自由度的非线性系统内,求解该方程也是十分困难的。于是,人们退而求其次,试图得到上述相空间中大气状态 PDF 的一阶矩(平均值)和二阶中心矩(方差)随时间的演变。尽管如此,对天气预报问题来说,求解一阶矩和二阶矩的时间演变方程,也是不可能的。因此,气象学家寻找了一种可以在实际应用中变通解决该问题的方法,这就是集合预报,具体做法为:通过一定的数学(概率统计)方法,获得在一定初值误差范围内,具有某种概率密度函数分布特征的初值集合,其中每个初值都有可能代表大气的真实状况。然后用数值模式对每个初值积分,从而得到一组预报结果的集合,再由这一组集合推断大气状态的 PDF 随时间的演变,这种方法被称为集合数值预报,简称集合预报(Molteni et al,1996;Zhu,2005;李泽椿等,2014)。可见,集合预报初始场的形成是概率密度函数的近似。从几乎海量的集合预报产品中提取有用的信息,形成直观快捷的图形图像、数据等产品,更是需要用到气象统计方法。当然,要想使用好这些集合预报产品,必须掌握相应的概率统计知识。

对罕见的极端天气和气候事件,数值模式的预报能力是很弱的。即使一个模式有能力预报它,也至少存在以下难点:一是有多大把握确定所预报的天气是极端事件?二是其具体的定时、定量、定点预报可靠吗?杜均等(2014)介绍了集合预报和气候资料相结合的“集合异常预报法”,并通过北京 2012 年 7 月 21 日(7·21)特大暴雨事件揭示出“集合异常预报法”和集合预报可以提供比单一模式预报更可靠和更准确的信息,从而有效地缓解上述两大难点。由于大气系统是复杂的非线性系统,加上初值和模式等本身无法避免的一些不确定性,天气预报从单一值的确定论向多值的概率论转变,不但符合气象科学的实际,也是更好地服务社会之必需(杜钧等,2010a)。集合预报产品的释用,应该更重视其提供的天气要素和事件的概率分布方面的信息,而不是

简单地只看其集合平均(薛纪善,2007)。集合预报的主要目的是定量描述预报的不确定性,即给出预报的概率密度函数,而不是给出一个更精确的单一值确定预报(杜钧等,2010c)。这些都说明,集合预报的基础是概率统计。

以集合预报为基础做出的概率预报,如何让公众接受并准确理解,是一个需要认真对待的问题。比如,“明天北京降水的概率为10%”意味着什么?概率统计上的准确解释为:如果对北京的明天做100次预报,其中有10次会出现降水。其实,概率是一个数学名词,其直观意思是指一件事情发生的可能性。北京的降水概率不是指降水的时间或区域,而是对降水是否出现这件事本身可能性的估计。

这种概率预报有什么意义呢?事实上,气象部门通过概率预报把天气预报的不确定性信息提供给用户,用户可以结合自己的实际情况,分析花费损失比,进行正确的决策。如果只是有、无危险天气出现的二分类预报,那么用户的最优决策除了要考虑预报的准确率以外,还需要考虑空报经济损失(或者收益)和漏报经济损失(或者收益)的比值。这个比值对不同的用户是不同的,不同的用户因此可以做出适合自己的相应决策(施能,2002;2009)。比如,对于风的预报,有两种预报:(1)明天风力2~4级(确定性预报);(2)明天风力5级以下的概率为70%,风力5级及以上的概率为30%。对于用温室大棚种植蔬菜的农民来说,当得到风力为2~4级的确定性预报信息时,根据往常的经验,这样的风力对他们的温室大棚不会产生危害,就不会采取防护措施;而当他们得到风力5级及以上的概率为30%的预报信息时,就会根据花费损失比,考虑是否采取措施加固其温室大棚,以防止温室大棚被大风刮坏。

可以预见,以集合预报为基础的概率预报必将成为未来的趋势。概率预报产品的推广和应用,离不开对概率统计知识进行广泛的科普和培训。

2.3 参数化方案

尽管数值模式的水平和垂直分辨率都在不断提高,但仍然有很多重要的过程和运动尺度,在目前的和将来的模式中都无法被显示和分辨,它们包括从几厘米到模式网格距大小的湍流运动与云的发展,以及发生在分子尺度的过程,比如凝结、蒸发、摩擦和辐射等。这些无法被数值模式显式分辨的过程,被称为“次网格尺度过程”。这些过程在数值模式中,

往往是根据观测或试验数据,采用统计方法,拟合出网格尺度的参数,这就是参数化方案(丑纪范,1986;沈桐立等,2003)。实际上,参数化方案对可分辨的与不可分辨的运动并没有进行尺度分离。参数化方案无法考虑可分辨和不可分辨气流的非线性相互作用,因此是造成数值模式系统性误差的一个重要原因。采用随机动力参数化方法可能是解决该问题的一种途径。

2.4 数值模式的统计释用

最近30年以来,气象数值模式的水平取得了长足进步,目前3天内的数值预报形势场已经超过经验预报,5~6天以内的数值预报形势场也已达达到可用水平。而人们最关注的气象要素(温度、湿度、降水量、风)和灾害性天气(如台风、暴雨、浓雾、大雪、沙尘暴等),数值模式或者效果不好,或者根本无法直接预报,因此对气象要素和灾害性天气预报来说,需要对数值预报产品进行解释应用,其中应用最广泛、最主要的方法就是统计释用。

在我国的中短期天气预报中,要素预报主要靠数值预报产品的统计释用(谢炯光等,2003;刘还珠等,2004;赵声蓉等,2012;曾晓青等,2013)。数值预报产品统计释用有多种方法,比如完全预报法(PP)、模式输出统计法(MOS)、卡尔曼滤波法、相似预报法、人工神经网络、支持向量法(SVM)、偏最小二乘等(矫梅燕,2010)。统计降尺度,是把数值模式预报较好的大尺度信息通过统计方法,得到局地的小尺度信息,从而实现从大尺度到小尺度的“降尺度”,也可以归为数值预报产品的统计释用。统计降尺度需要综合应用多种统计方法,比如EOF、多元回归、典型相关(CCA)、聚类分析、交叉验证等(Rasmus et al,2008)。

特别需要指出的是,EOF可以对时空信息交织在一起的气象信息进行时空分离,而分离开的分量又具有相互正交的优良性质,从某种程度上可以克服预报时气象因子不独立的问题,大大压缩了因子维数并基本保留了气象因子的初始信息,因此在数值预报产品统计释用中具有广泛应用(Hannachi et al,2007)。熊开国等(2012)提出将模式误差的直接相似订正问题转化成模式误差主分量的相似预报问题,做法为对模式误差距平场进行EOF分解,取前1~9个空间模态作为“可预报”模态进行最优多因子动态配置相似预报。于杰等(2014)将WRF中

尺度数值模式模拟结果作为实况资料集,应用 EOF 方法对该资料集进行诊断,探讨该方法用于暴雨 β 中尺度系统的可行性,是数值预报产品统计释用的一种新尝试。基于统计方法的对数值模式产品的降尺度在气候预测中也具有广泛应用 (Chen et al, 2012; Guo et al, 2014)。龚志强等(2015)采用 EOF 分析方法,得到东亚夏季降水的年代际变化分量,基于海温关键区的动力-统计相结合的预报方案,嵌入了模式对东亚夏季降水季节预报结果中的年代际变化信息。

2.5 延伸期预报

延伸期预报是指 10~30 d 这个时间段的预报。与初始条件相关的可预报性称为“第一类可预报性”,而与缓慢变化的边界条件相联系的可预报性称为“第二类可预报性”(Kalnay, 2001)。两周以内的预报可以归为第一类可预报性问题,主要由初始条件控制。月以上的预报可以归为第二类可预报性问题,主要是大气向上下边界(太阳辐射、海温、土壤湿度、冰雪等)强迫的适应。从目前的研究来看,延伸期预报还存在很大的困难,其原因在于其预报时效超越了逐日天气预报时间的理论上限(两周左右),而预报对象的时间尺度又小于气候预测的月、季时间尺度(朱玉祥等, 2013)。延伸期预报的方法大致可以分为纯统计方法、纯动力方法、动力-统计相结合的方法,而统计方法与数值模式相结合的动力-统计方法更是引起了广泛关注。丑纪范等(2010)针对大气系统的混沌特性,以数值模式为基础,阐述了延伸期可预报分量的提取方法。郑志海等(2012)发展了一种基于延伸期可预报性的集合预报新方法,该方法以数值预报模式为平台,对可预报分量采用多个模式误差订正方案,从考虑模式不确定性的角度进行集合,而对随机分量则利用历史资料从气候概率的角度给出集合概率分布,避免模式误差对随机概率分量概率分布的影响,试验结果表明该方法显示出潜在的业务应用前景。郑志海等(2013)采用 EOF 方法压缩数值模式的自由度,并且用方差分析方法来计算 EOF 分量的可预报性,同时结合历史资料,利用相似-动力方法对可预报分量的预报误差进行订正,达到减小模式误差和从统计角度考虑随机分量对可预报分量影响的目的,结果表明,该方法能有效提高数值模式对可预报分量的预报技巧。Feng 等(2013)以 2009 年初长江中下游的连续连阴

雨过程为例,采用 EOF 和方差分析方法,从 NCEP-DOE 再分析资料中分析了延伸期天气预报的稳定分量。王启光等(2014)借鉴 CNOP 相关算法,形成了在数值模式中提取延伸期尺度的可预报分量的实用方法和预报技术,数值试验结果表明该方法能有效地提高数值模式中 10~30 d 延伸期预报的预报技巧。

可见,基于数值模式预报结果,结合统计方法做延伸期预报是一条值得深入探索的可行途径。

2.6 数值模式的检验

由于数值模式产品存在偏差,因此在使用前必须要对其进行检验评估,其中统计检验具有客观定量的特点,应用最为广泛(王雨, 2006; 王雨等, 2007; 潘留杰等, 2014; 赖芬芬, 2015)。具体统计检验方法可以参考 Ian 等(2012)的文献。

当然,分析数值模式产品,做统计推断的时候,更需要高度重视统计检验。因为气象变量是在时间上无始无终一直发展变化的变量,而我们分析的对象,即我们得到的数值模式资料只能是某一时段(比如过去 50 年)的离散资料。根据概率统计的术语,我们得到的资料可以称为样本,是有限的,而气象变量的总体是无限的。当从有限的样本推断无限总体的性质时,必须进行显著性检验。使用统计方法分析气象问题,做出统计推断时,也必须要进行统计显著性检验。比如,在做数值模拟时,需要检验控制试验和敏感性试验的差别是否显著(朱玉祥等, 2009)。

统计检验的一般步骤可参见相关文献(黄嘉佑, 2004; 魏凤英, 2007),黄嘉佑等(2015)介绍了平均值检验、方差检验、相关系数检验、变量的分布检验、频率检验、趋势检验、突变检验的常用方法,这些方法大都假设变量遵从正态分布。在不少文献中,统计检验存在不恰当应用和文字表述,关于这方面的分析,可以参见相关文献(Falk, 1986; Nicholls, 2001; 黄嘉佑, 2005; 施能, 2009; Ambaum, 2010)。

传统的统计检验方法,一般都假定研究对象服从正态分布,但很多气象变量,比如现在广泛关注的极端天气气候事件,往往不符合正态分布(Easterling, 2000; Meehl et al, 2000; 丁裕国, 2007; 丁裕国等, 2009; Qian et al, 2015)。另外,气象时间序列的不独立性、空间距离的非均匀性、时间尺度的不确定性等问题,都是传统统计方法在检验时无法解决的,而自助法(bootstrap)在解决这些问题时表现出了

很好的能力(Mudelsee, 2014)。自助法的核心思想是,我们使用的气象资料(可以看作一次抽样),与把这些资料打乱顺序进行随机抽样相比较,如果有显著差异,则认为我们从观测资料得到的结论是可靠的。在目前计算机能力高速发展的今天,可以采用把样本(气象资料)与多次(1000 或 10000 次)随机抽样相比较的方法进行检验。

还有一种被称为蒙特卡罗检验的随机检验技术,其思想为利用随机数发生器多次(1000 或 10000 次),产生出与分析的问题所使用的资料具有相同概率分布、相同序列长度和空间分布的资料序列,通过对随机生成的多个时间资料序列进行计算得到概率分布,确定出小概率事件落入的否定域,对检验对象是否落在否定域进行鉴别,做出接受或拒绝原假设的判别。施能等(1997)指出了气象场相关分析及合成分析中的问题,并且介绍了解决这种问题的蒙特卡罗检验方法。周永宏等(1999)应用蒙特卡罗方法,通过大量的数值计算,获得了一个容量大、稳定性好的相关系数临界值表,这种方法付出了大量的计算代价,但是摆脱了相关系数检验时所需要满足的变量正态分布的假定。Zhang 等(2004)使用蒙特卡罗模拟,比较了极值趋势检测的几种方法。施能等(2004)提出了检验二维合成风场差异是否显著的蒙特卡罗检验方法,施能(2009)的文献中给出了该检验方法的源程序。EOF 分析和 SVD 分析也可采用此种方法进行检验(魏凤英, 2007; 施能, 2009)。

相关分析在气象科研和业务中具有广泛应用。在做气候变化研究时,也经常需要对气候模式资料做相关分析。在全球变暖的背景下,很多气象变量往往存在趋势。对于带有趋势的两个时间序列计算相关时,由于趋势的存在,样本并不相互独立,不再满足传统的 t 检验条件。施能等(2007)对该问题进行了数值试验,结果表明:变量带有性质相反的趋势变化,会使这两个变量之间的相关系数减小(正相关的数值被减小,负相关被夸大);变量带有性质相同的趋势变化,会使这两个变量之间的相关系数增加(正相关被夸大,负相关数值被减小)。该研究结论对正确合理地进行相关分析具有重要意义。

3 贝叶斯统计的应用

虽然贝叶斯统计在分析数值模式资料中已经有了一些应用,但鉴于贝叶斯统计的特点,可以预计在

未来的气象科研和业务中,仍然会有很大的发展和提高潜力。因此,独立出来重点阐述。

贝叶斯统计的基础是贝叶斯理论(Berger, 2000; 2004; 茆诗松等, 2012)。贝叶斯统计学派与经典统计学派在很多问题上都有分歧,但它们最根本的分歧点在于:第一,是否利用先验信息。对于很多问题来说,都存在先验信息可以利用(比如,对于气象预报来说,历史资料可以看作先验信息),贝叶斯统计学派认为利用这些先验信息不仅可以减少样本容量,而且在很多情况还可以提高估计精度;而经典统计学派忽略了这些信息。第二,是否将参数看成随机变量。贝叶斯统计学派的最基本观点是:任一未知量都可以看成随机变量,可以用一个概率分布去描述,这个分布就是先验分布。因为任一未知量都具有不确定性,而在表述不确定性时,概率与概率分布是最好的描述语言;相反,经典统计学派却把未知量简单看成一个未知参数,来对它进行统计推断。

贝叶斯统计分析具有两个基本功能:(1)信息的有效组合;(2)量化和处理不确定性。对于天气和气候科学来说,面对的是海量数据以及复杂的非线性过程问题,需要对各种信息进行有效组合,并且量化其不确定性,因此贝叶斯模拟和分析是一个应用的基本策略(Berliner et al, 1999)。数值模式是天气预报和气候预测的基础,天气预报和气候预测如果只是单一值的决定论式预报,那么它是不完备的,预报的不确定性也应该作为预报不可分割的一部分定量地告知用户和公众(杜钧等, 2010b)。预报与不确定性是共生共存的,一个预报如果没有定量地估计它的不确定性,那么该预报是不完全的(王东海等, 2011)。在地球气候系统模式研究中,实际上是追求气候科学和不确定性研究方面的进步(周天军等, 2014)。可见,量化和减小不确定性,是天气预报和气候预测的重要任务,因此适合应用贝叶斯统计方法。

贝叶斯统计由于结合了数据的信息与参数的先验信息,不断通过样本数据更新先前的认知,因此相比经典统计具有明显的优势。对于天气预报和气候预测问题来说,具有长期的历史观测资料,并且不断得到新的信息,因此可以根据新信息不断改进预报预测结果,贝叶斯统计能够符合天气预报和气候预测的这一特点。丑纪范(1974)从微分方程只是近似地描述了大气中的物理过程的观点出发,提出了在数值天气预报中使用历史资料来考虑场演变的时间

连续性问题,其实就是贝叶斯统计思想在数值预报中应用的一个例子。Guan 等(2015)采用全球集合预报系统的回报资料改进数值预报的后处理,也是贝叶斯统计思想在数值模式产品释用中应用的案例。

贝叶斯统计在数值模式产品中应用的例子还有很多。Elsner 等(2004)用分级贝叶斯模型研究美国登陆飓风与 ENSO 和 NAO 之间的关系。Min 等(2006;2007)采用贝叶斯方法,对全球和区域的地面气温进行检测归因,都取得了较好的效果。DelSole(2007)提出了一种基于贝叶斯回归的方法,并且使用该方法对欧洲多模式预测系统(DEMETER)进行温度的季节预测,其结果表明,与传统回归模型相比,该方法在几个地区的预报效果有明显提高。Lean 等(2007)提出了一种基于数值预报的集成员,可以预报定量降水的概率密度函数的贝叶斯平均方法,该方法可以对未来的降水预报出完整的概率分布。资料同化中的三维和四维变分同化也使用了贝叶斯理论(邹晓蕾,2009)。陈朝平等(2010)基于贝叶斯方法,对四川暴雨的集合概率预报产品进行了释用,在一定程度上消除了空报。为了克服单个大气环流模式对未来预估的不确定性,Malaak 等(2010)提出了一种分层贝叶斯模型,对多个大气环流模式的未来预估结果进行集成,对北极海洋地区累积海冰的演化情况进行评估。Chu 等(2010)使用贝叶斯方法做台风活动的季节预测。Zhao 等(2010)发展了一种可以检测不服从正态分布的极端天气气候事件(比如台风、强降水和高温热浪)突变点的分层贝叶斯方法。Chu 等(2011)对极端气候事件使用贝叶斯分析进行了全面综述。Wilks(2011)对贝叶斯理论在气象中的应用做了阐述。陈法敬等(2011b)基于单一数值预报产品的贝叶斯统计处理技术-贝叶斯处理器,对 NCEP 集合预报各成员进行建模,获得了各成员的贝叶斯概率预报,该方法可以将集合预报的不确定性量化为一个集成贝叶斯概率预报,从而实现集合预报的概率化。陈法敬等(2011a)基于贝叶斯统计理论,以中国国家气象中心、欧洲中期天气预报中心、美国国家环境预测中心集合预报中的控制资料作为确定性预报样本,采用亚高斯贝叶斯预报处理器,将确定性预报转化为能定量地表达各控制预报的不确定性的概率预报。李芳(2012)采用包括等权重(EE)、校正 EE、多元回归、贝叶斯统计的 4 种方案,对 5 个海-气耦合模式

的后报降水资料进行多模式集合,结果表明贝叶斯方案在华南最优,并且对中国东部总体比较不同方案的预报,发现贝叶斯方案有最好的预报效果。Michael 等(2012)对采用贝叶斯平均模型前后的多模式集合的性能进行了评估,结果表明贝叶斯平均模型对概率预报技巧有改进。Madadgar 等(2013)在贝叶斯统计框架内提出了一种可以做季节干旱概率预测的方法。杜均等(2014)采用集合预报与气候资料相结合,对北京 2012 年 7 月 21 日特大暴雨的预报,也是贝叶斯统计思想的应用。Xu 等(2014)提出一种新的两步贝叶斯随机动力模型,来处理水利发电实时业务系统中来自定量降水预报的不确定性。Otto(2014)给出了 Heidke 技巧评分的概率推导,该方法本质上是贝叶斯的,因为它把先验概率更新到后验概率中。

贝叶斯统计在数值模式分析和释用中,虽然已经取得了很多成果,但其特点决定了在未来气象科研和业务的发展中,贝叶斯统计的应用空间将更加广阔。

4 统计方法在数值模式中应用的趋势展望

随着数值模式水平的不断进步,以及人们对极端天气气候事件的高度关注,气象统计在数值模式中的应用面临很多新问题,我们认为其中下面几个问题值得深入研究:

(1) 气象变量的概率分布,特别是不符合正态分布的气象变量的概率分布的理论基础。概率分布是应用各种统计方法的基础,但对于不少气象变量,其概率分布形成的理论基础并不完善,值得深入研究。物理学上的熵理论可能是研究该问题的有力工具。

(2) 准确估计不同气象变量之间的联合概率密度函数。比如,资料同化中需要对观测资料误差和模式误差进行估计,由于观测资料之间和模式网格点之间存在复杂的相互关系,准确地估计出其联合概率密度函数,对于提高资料同化水平具有重要意义。

(3) 随机动力学,或者说如何在数值模式中体现气象变量的随机变化。正如 Lorenz 在 1975 年所言:“我相信,最终的气候模式将是随机的,即随机数将出现在方程时间导数的某些地方”。加强对利用

概率密度函数描述气象变量空间分布的非均匀性和时间变化的研究,可能有助于在数值模式中体现气象变量的随机变化。

(4) 在数值预报产品统计释用中克服不同气象变量的复相关问题。由于气象变量之间存在各种复杂的相互关系(不独立),导致在做数值预报产品的统计释用时,在历史时段常常会拟合得很好,甚至会出现过拟合现象,而用于对未来时段的预测时泛化能力不足。主成分(EOF)分析可能有助于解决该问题。

(5) 如何从海量的数值预报产品中提取有价值的信息?可以预见,未来全球各大业务和研究中心的多个数值模式,会产生不计其数的分析、再分析、预报预测、预估等各种数值模式产品,这些产品将构成海量的庞大数据。如何从这些数据中提取对我们的气象服务有用的信息,并且量化其中的不确定性,数据挖掘(Data Mining)技术可能是一种有效的手段。数据挖掘中的分类、估计、预测、聚类、描述和可视化等方法,都离不开统计(Fernandez, 2010)。

(6) 贝叶斯统计的思想,如何在基于数值模式的天气预报和气候预测中得到更深入的应用,特别是针对不同的气象变量和不同的时间尺度,如何合理地选取先验概率和似然函数值得深入研究。

5 结 论

综上所述,统计方法是数值模式的重要组成部分。在强调数值模式的巨大价值的同时,绝不应该忽视统计方法的重要性。近年来,我国的数值模式研究取得了长足进步,但进一步提高模式性能的难度越来越大。在重视动力方法的同时,也应该高度重视统计方法的研究,动力和统计的深入结合,可能是提高天气预报和气候预测水平的有效手段。在应用统计方法的时候,需要高度重视检验问题,特别是对于极端天气气候事件等不符合正态分布的气象变量,不能采取传统的基于正态分布的检验方法。贝叶斯统计的特点决定了其在天气预报和气候预测中具有巨大的应用潜力。

要想提高数值预报水平,需要深入研究气象统计理论与方法。数值预报越发展,对概率统计理论与应用的要求就越高,气象统计是未来值得深入研究的一个领域。

参考文献

- 陈朝平,冯汉中,陈静. 2010. 基于贝叶斯方法的四川暴雨集合概率预报产品释用. 气象, 36(5): 32-39.
- 陈法敬,矫梅燕,陈静. 2011a. 一种温度集合预报产品释用方法的初步研究. 气象, 37(1): 14-20.
- 陈法敬,矫梅燕,陈静. 2011b. 亚高斯贝叶斯预报处理器及其初步试验. 气象学报, 69(5): 872-882.
- 丑纪范. 1974. 天气数值预报中使用过去资料的问题. 中国科学 数学辑, 4(6): 635-644.
- 丑纪范. 1986. 为什么要动力-统计相结合-兼论如何结合. 高原气象, 5(4): 367-372.
- 丑纪范,郑志海,孙树鹏. 2010. 10—30d 延伸期数值天气预报的策略思考——直面混沌. 气象科学, 30(5): 569-573.
- 丁裕国. 2007. 气候概率分布理论的新内涵及其展望. 沙漠与绿洲气象, 1(2): 1-5.
- 丁裕国,江志红. 1998. 气象数据时间序列信号处理. 北京: 气象出版社, 285.
- 丁裕国,申红艳,江志红,等. 2009. 气候概率分布理论及其应用新进展. 气象科技, 37(3): 257-262.
- 杜钧,陈静. 2010a. 天气预报的公众评价与发布形式的变革. 气象, 36(1): 1-6.
- 杜钧,陈静. 2010b. 单一值预报向概率预报转变的基础: 谈谈集合预报及其带来的变革. 气象, 36(11): 1-11.
- 杜钧,邓国. 2010c. 单一值预报向概率预报转变的价值: 谈谈概率预报的检验和应用. 气象, 36(12): 10-18.
- 杜钧,Grumm R H,邓国. 2014. 预报异常极端高影响天气的“集合异常预报法”: 以北京 2012 年 7 月 21 日特大暴雨为例. 大气科学, 38(4): 685-699.
- 龚志强,赵俊虎,封国林,等. 2015. 基于年代际突变分量的东亚夏季降水动力-统计预报方案研究. 中国科学 地球科学, 45(2): 236-252.
- 黄嘉佑. 1990. 气象统计分析与预报方法. 北京: 气象出版社, 358.
- 黄嘉佑. 2004. 气象统计分析与预报方法(第3版). 北京: 气象出版社, 298.
- 黄嘉佑. 2005. 气象中使用统计检验的几个问题. 气象, 31(7): 3-5.
- 黄嘉佑,李庆祥. 2015. 气象数据统计分析方法. 北京: 气象出版社, 506.
- 矫梅燕. 2010. 现代数值预报业务. 北京: 气象出版社, 206.
- 赖芬芬. 2015. 2015 年 3—5 月 T639、ECMWF 及日本中期预报性能检验. 气象, 41(8): 1036-1041.
- 李芳. 2012. 基于多模式集合方案的中国东部夏季降水概率季度预测. 气象学报, 70(2): 183-191.
- 李泽椿,毕宝贵,金荣花,等. 2014. 近 10 年中国现代天气预报的发展与应用. 气象学报, 72(6): 1069-1078.
- 刘还珠,赵声蓉,陆志善,等. 2004. 国家气象中心气象要素的客观预报—MOS 系统. 应用气象学报, 15(2): 181-191.
- 马开玉,丁裕国,屠其璞,等. 1993. 气候统计原理与方法. 北京: 气象出版社, 518.
- 茆诗松,汤银才. 2012. 贝叶斯统计. 北京: 中国统计出版社, 302.
- 穆穆,陈博宇,周非凡,等. 2011. 气象预报的方法与不确定性. 气象,

- 37(1):1-13.
- 潘留杰,张宏芳,王建鹏. 2014. 数值天气预报检验方法研究进展. 地球科学进展, 29(3):327-335.
- Eugenia Kalnay, 著. 蒲朝霞,杨福全,邓北胜,等. 2005. 大气模式、资料同化和可预报性. 北京:气象出版社,300.
- 沈桐立,田永祥,葛孝贞,等. 2003. 数值天气预报. 北京:气象出版社,471.
- 施能. 2002. 气象科研与预报中的多元分析方法. 北京:气象出版社,192.
- 施能. 2009. 气象统计预报. 北京:气象出版社,287.
- 施能,顾骏强,黄先香,等. 2004. 合成风场的统计检验和蒙特卡洛检验. 大气科学, 28(6):950-956.
- 施能,顾骏强,封国林. 2007. 论带有趋势的变量的相关:数值试验. 数学的实践与认识, 37(8):98-104.
- 施能,魏凤英,封国林. 1997. 气象场相关分析及合成分析中的蒙特卡洛检验. 南京气象学院学报, 20(3):355-359.
- 施能,章爱国,余锦华. 2009. 气象学中使用统计检验的几个重要笔记. 气象科学, 21(4):522-524.
- 王东海,杜钧,柳崇健. 2011. 正确认识和对待天气气候预报的不确定性. 气象, 37(4):385-391.
- 王启光,丑纪范,封国林. 2014. 数值模式延伸期可预报分量提取及预报技术研究. 中国科学 地球科学, 44(2):343-354.
- 王雨. 2006. 2004 年主汛期各数值预报模式定量降水预报评估. 应用气象学报, 17(3):316-324.
- 王雨,闫之辉. 2007. 降水检验方案变化对降水检验评估效果的影响分析. 气象, 33(12):53-61.
- 魏凤英. 2006. 气候统计诊断与预测方法研究进展——纪念中国气象科学研究院成立 50 周年. 应用气象学报, 17(6):736-742.
- 魏凤英. 2007. 现代气候统计诊断与预测技术. 北京:气象出版社,296.
- 魏淑秋. 1985. 农业气象统计. 福州:福建科学技术出版社,377.
- 吴洪宝,吴蕾. 2005. 气候变率诊断和预测方法. 北京:气象出版社,371.
- 谢炯光,曾琮,纪忠萍. 2003. 中国近 30 年来气象统计预报进展. 气象科技, 31(2):67-83.
- 熊开国,赵俊虎,封国林,等. 2012. 汛期降水相似动力预报——模式误差主分量相似预报方法. 物理学报, 61(14):149204(1-11).
- 薛纪善. 2007. 和预报员谈数值预报. 气象, 33(8):3-11.
- 么枕生,丁裕国. 1990. 气候统计. 北京:气象出版社,954.
- 于杰,张继权,张铭. 2014. EOF 分析用于 β 中尺度暴雨系统的探索. 大气科学, 38(4):795-803.
- 曾晓青,赵声蓉,段云霞. 2013. 基于 MOS 方法的风向预测方案对比研究. 气象与环境学报, 29(6):140-144.
- 赵声蓉,赵翠光,赵瑞霞,等. 2012. 我国精细化客观气象要素预报进展. 气象科技进展, 2(5):11-20.
- 郑志海,封国林,黄建平,等. 2012. 基于延伸期可预报性的集合预报方法和数值试验. 物理学报, 61(19):199203(1-8).
- 郑志海,黄建平,封国林,等. 2013. 延伸期可预报分量的预报方案和策略. 中国科学 地球科学, 43(4):594-605.
- 周家斌,黄嘉佑. 1997. 近年来中国统计气象学的新进展. 气象学报, 55(3):297-305.
- 周天军,邹立维,吴波,等. 2014. 中国地球气候系统模式研究进展:CMIP 计划实施近 20 年回顾. 气象学报, 72(5):892-907.
- 周永宏,郑大伟. 1999. 相关估计显著水平的 Monte Carlo 模拟检验. 测绘学报, 28(4):313-318.
- 朱国富. 2015a. 数值天气预报中分析同化基本方法的历史发展脉络和评述. 气象, 41(8):986-996.
- 朱国富. 2015b. 理解大气资料同化的内在逻辑和若干共性特征. 气象, 41(8):997-1006.
- 朱玉祥,丁一汇,刘海文. 2009. 青藏高原冬季积雪影响我国夏季降水的模拟研究. 大气科学, 33(5):903-915.
- 朱玉祥,俞小鼎,赵亮,等. 2013. 10—30 天延伸期预报及其策略思考. 沙漠与绿洲气象, 7(4):38-44.
- 邹晓蕾. 2009. 资料同化理论和应用. 北京:气象出版社,120.
- Ambaum M H P. 2010. Significance tests in climate science. J Climate, 23(22): 5927-5932.
- Berger J O. 2000. Bayesian analysis: A look at today and thoughts of tomorrow. J Amer Statist Assoc, 95(452): 1269-1276.
- Berger J O. 2004. Statistical Decision Theory and Bayesian Analysis. London: Springer-Verlag,617.
- Berliner L M, Royle J A, Wikle C K, et al. 1999. Bayesian Methods in the Atmospheric Sciences// Bernardo J M, Berger J O, Dawid A P et al. Bayesian Statistics London: Oxford University Press, 83-100.
- Chen Huopo, Sun Jianqi, Wang Huijun. 2012. A statistical downscaling model for forecasting summer rainfall in China from DEMETER Hindcast Datasets. Wea Forecasting, 27: 608-628.
- Chu P S, Zhao X, Ho C H, et al. 2010. Bayesian forecasting of seasonal typhoon activity: A track-pattern-oriented categorization approach. J Climate, 23(24): 6654-6668.
- Chu P S, Zhao X. 2011. Bayesian analysis for extreme climatic events: A Review Atmos Res, 102: 243-262.
- DelSole Timothy. 2007. A bayesian framework for multimodel regression. J Climate, 20(12): 2810-2826.
- Easterlling D R. 2000. Observed variability and trends in extreme climate events: A brief review. Bull Amer Meteor Soc, 81(3): 417-425.
- Elsner J B, Jagger T H. 2004. A hierarchical Bayesian approach to seasonal hurricane modeling. J Climate, 17(14): 2813-2827.
- Falk R. 1986. Misconceptions of statistical significance. J Struct Learn, 9: 83-96.
- Feng G L, Sun S P, Zhao J H, et al. 2013. Analysis of stable components for extended-range (10—30 days) weather forecast: A case study of continuous overcast-rainy process in early 2009 over the mid-lower reaches of the Yangtze River. Science China Earth Sciences, 56: 1576-1587.
- Fernandez George. 2010. Statistical Data Mining Using SAS Applications (Second Edition). London: CRC Press, 442.
- Guan Hong, Cui Bo, Zhu Yuejian. 2015. Improvement of statistical postprocessing using GEFS reforecast information. Wea Forecasting, 30: 841-854.
- Guo Yan, Li Jianping, Li Yun. 2014. Seasonal forecasting of North

- China summer rainfall using a statistical downscaling model. *J Appl Meteor Climat*, 53(7): 1739-1749.
- Hannachi A, Jolliffe I T, D B. 2007. Stephenson, Empirical orthogonal functions and related techniques in atmospheric science: A review. *Inter J Climat*, 27(9): 1119-1152.
- Ian T J, Stephenson D B. 2012. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. London: Wiley Press, 240.
- Kalnay E. 2001. *Numerical Weather Forecasting and Predictability*. London: Cambridge University Press, 341.
- Lean Sloughter J, Mc, Adrian E R, Tilmann G, et al. 2007. Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging. *Mon Wea Rev*, 135(9): 3209-3220.
- Madadgar S, Moradkhani H. 2013. A Bayesian framework for probabilistic seasonal drought forecasting. *J Hydrometeor*, 14: 1685-1705.
- Malaak K, Maksimovich E, Michelangeli P A, et al. 2010. Multi-model Combination by a Bayesian Hierarchical Model: Assessment of Ice Accumulation over the Oceanic Arctic Region. *J Climate*, 23(20): 5421-5436.
- Meehl G A, Karl T R, Easterling D R, et al. 2000. An introduction to trends in extreme weather and climate events: Observations, socioeconomic impacts, terrestrial ecological impacts, and model projections. *Bull Amer Meteor Soc*, 81(3): 413-416.
- Michael J E, Brian A C, Joseph J C. 2012. Impact of Bias-correction type and conditional training on Bayesian model averaging over the Northeast United States. *Wea Forecasting*, 27: 1449-1469.
- Min S K, Hense A. 2006. A Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models. *Geophys Res Lett*, 33, L08708. DOI: 10.1029/2006GL025779.
- Min S K, Hense A. 2007. A Bayesian assessment of climate change using multimodel ensembles. Part II: Regional and Seasonal mean surface temperature. *J Climate*, 20(12): 2769-2790.
- Molteni F, Palmer T N, Buizza R, et al. 1996. The ECMWF ensemble prediction system: Methodology and verification. *Quart J Roy Met Soc*, 122: 73-121.
- Mu Mu, Zhou Feifan, Wang Hongli. 2009. A method for identifying the sensitive areas in targeted observations for tropical cyclone prediction: Conditional nonlinear optimal perturbation. *Mon Wea Rev*, 137: 1623-1639.
- Mudelsee M. 2014. *Climate Time Series Analysis Classical Statistical and Bootstrap Methods*. London: Springer Press, 474.
- Nicholls N. 2001. The insignificance of significance testing. *Bull Amer Meteor Soc*, 82: 981-986.
- Otto Hyvärinen. 2014. A probabilistic derivation of Heidke skill score. *Wea Forecasting*, 29: 177-181.
- Palmer T, Renate H. 2006. *Predictability of Weather and Climate*. Cambridge University Press: New York, 702.
- Qian Cheng, Zhou Wen, Fong Soi Kun, et al. 2015. Two Approaches for Statistical Prediction of Non-Gaussian Climate Extremes: A Case Study of Macao Hot Extremes during 1912-2012. *J Climate*, 28(2): 623-636.
- Rasmus E B, Inger H B, Chen Deliang. 2008. *Empirical-Statistical Downscaling*. London: World Scientific Publishing Company, 215.
- von Storch H, Zwiers F W. 1999. *Statistical Analysis in Climate Research*. Cambridge: Cambridge University Press, 484.
- Wilks D S. 2011. *Statistical Methods in the Atmospheric Sciences (3rd edn)*. Amsterdam: Academic Press, 676.
- Xu W, Zhang C, Peng Y, et al. 2014. A two stage Bayesian stochastic optimization model for cascaded hydropower systems considering varying uncertainty of flow forecasts. *Water Resour Res*, 50: 9267-9286.
- Zhang Xuebin, Zwiers F W, Li Guilong. 2004. Monte Carlo experiments on the detection of trends in extreme values. *J Climate*, 17(10): 1945-1952.
- Zhao Xin, Chu P S. 2010. Bayesian change point analysis for extreme events (Typhoons, Heavy Rainfall, and Heat Waves): An RJMCMC Approach. *J Climate*, 23(5): 1034-1046.
- Zhu Yuejian. 2005. Ensemble forecast: A new approach to uncertainty and predictability. *Adv Atmos Sci*, 22(6): 781-788.
- Zwiers F W, von Storch H. 2004. On the role of statistics in climate research. *Int J Climatol*, 24: 665-680.