

基于 Ward 聚类法的中国业务 集合预报系统的产品开发

金荣花 田伟红 矫梅燕

(国家气象中心,北京 100081)

提 要: 为开发中国新一代业务集合预报系统的聚类产品,系统地介绍了集合预报技术先进国家的集合预报聚类分析方法,并采用 Ward 聚类法开发了中国集合预报系统的聚类产品。针对重大天气过程初步分析了中国集合预报系统的聚类产品的业务预报能力。结果表明,Ward 聚类法以凝练预报信息和给出发生概率的形式,有效地划分出最有可能出现的环流形势类型,为预报员提供了有价值的预报参考信息,也为科学合理地使用集合预报产品提供方便。

关键词: Ward 聚类法 集合预报系统 聚类产品开发

Product Development for the Operational Ensemble Prediction System Based on Ward Analysis in China

Jin Ronghua Tian Weihong Jiao Meiyan

(National Meteorological Center, Beijing 100081)

Abstract: In order to develop the cluster products for Chinese new generation operational Ensemble Prediction System (EPS), the cluster analysis techniques for EPS used in foreign developed country are introduced, and the Ward analysis is applied to interpret the products of Chinese EPS. The case study for operational application capacity of Chinese EPS cluster products indicates that the Ward analysis can classify the circulation situation efficiently, and the cluster products are very helpful and convenient for the weather forecaster to make use of the EPS information.

Key Words: Ward analysis Chinese EPS cluster product development

资助项目:国家气象中心自筹资金课题《中国集合预报产品业务应用研究》(ZK2006-08)和《TIGGE 多中心数据的应用研究》(ZK2007-0102)共同资助

收稿日期:2007年6月12日; 修定稿日期:2007年10月17日

引言

在任何时候大气的真实状态是未知的,只能近似地描述。所以确切地说,天气预报问题应该提为在大气相空间中合适的概率密度函数随时间的演变^[1]。当误差的增长超出线性动力学范畴之外,集合预报是预报大气状态概率密度函数演变唯一可行的方法。集合预报代表数值预报未来发展的方向,尤其是进入 21 世纪 TIGGE 计划的实施,更是促进了集合预报的发展。

集合预报思想由 Epstein^[2] 和 Leith^[3] 提出后,经过了近 20 年的研究探索和数值预报试验,到 1990 年代初,集合数值预报进入了实际业务应用阶段。美国国家环境预报中心^[4-6] 和欧洲中期天气预报中心于 1992 年率先建立各自的中期集合数值预报业务系统,随后几年,英国、日本、南非、法国、加拿大等发达国家的气象部门也相继建立了各自的业务集合预报系统。经过近 15 年的业务应用,集合预报业务产品的种类、表现形式以及评价标准等的认识已经趋于统一,对集合预报产品的应用技术已经相当成熟。相比之下,中国开展集合数值预报研究和业务应用^[7-10],尤其在产品的解释应用方面,显得投入不足和技术薄弱。中国国家气象中心于 1995 年 5 月下旬开始以原有的动力延伸预报为基础,着手在 IBM/SP2 巨型并行计算机上开发建立一套中期集合数值预报系统,该系统以 T63L16 为基础,采用时间滞后平均法生成扰动初值,由 12 个成员组成集合预报系统,于 1996 年 9 月投入业务试验运行。此后经过了 3 次升级和改造,2005 年底,以 T213L31 为基础采用繁殖增长模法生成 15 个扰动初值的新一代集合预报系统。与原集合预报系统相比较,该系统的扰动初值生成技术以及各成员的发散度等,均有较大程度

的提高或已基本接近国际先进水平。然而,由于集合预报业务应用问题一直以来未能在我国业务部门引起足够重视,造成集合预报业务产品种类单一,解释应用能力与评价标准和方法等仍停留在 1990 年代中后期的水平上。

从业务预报的角度来看,要在一个有限的时间内,找出集合预报系统输出的多个预报场间的相同点和差异几乎是不可能的。因此,需要对这些预报场进行处理合成。聚类分析方法是实现集合预报产品信息浓缩和环流类型划分的一种简捷、有效的方法。在许多国家气象业务部门普遍使用集合预报聚类产品,然而,在我国新一代业务集合预报系统中,尚没有开发和使用集合预报聚类产品。为此,本文首先介绍了集合预报技术先进国家的聚类分析方法,通过不同方法的比较,结合我们业务应用的实际情况,进一步将 Ward 分析方法应用到中国业务集合预报系统的产品释用中,并给出了初步分析结果。

1 国外业务集合预报系统聚类分析方法

聚类分析方法^[6]是把集合预报中相似的成员合并成一类,同时给出该类出现的相对频率。特别对于多平衡态的大气状态,聚类法可以提供很清晰的具有几种典型平衡态的预报指导,使得聚类法更适合于经验不多的预报员。相似度的标准一般可用相关系数或均方差等度量,值得注意的是,根据选择相似度的标准不同,在同一个集合预报系统中,聚类得出的某一类形势的构成成员也许会不一样。国外常用的聚类分析法主要有 Ward 聚类法^[11]、管子法^[12] (Tubing clustering)、距平相关系数分簇法、动力模糊法、神经元聚类法和中央聚类法 (Central clustering) 等。下面简要介绍这些方法及其业务应用情况,便于今后在我国的业务集合预报系统中加以借

鉴使用。

ECMWF 开始采用 Ward^[11] 聚类法,后采用管子法^[12]对集合预报产品进行聚类。Ward 最小方差法^[13](或称 Ward 离差平方和法)聚类时首先使 n 个样本各自成一类,然后样本之间离差平方和最小的两个样本合并成一类,这时类的数目减至 $n-1$ 类,然后再合并其中的两类,使得类内离差平方和增加最小,直至所有的样本聚成一类为止,Ward 方法总是使得聚类导致的类内离差平方和增量最小。主要的计算公式如下:

第 K 个成员和第 J 个成员之间距离为:

$$D_{K,J} = \|X_K - X_J\|^2 / (1/n_K + 1/n_J) \quad (1)$$

第 L 个成员与 K, J 合并成一组的之间的距离为:

$$D_{L,JK}^2 = \frac{n_L + n_J}{n_L + n_J + n_K} D_{L,J}^2 + \frac{n_L + n_K}{n_L + n_J + n_K} D_{L,K}^2 - \frac{n_L}{n_L + n_J + n_K} D_{J,K}^2 \quad (2)$$

这种分类的思想^[11]来自方差分析,如果类分得好,同类样本的离差平方和应当小,类与类之间的离差平方和应当较大。因此,Ward 聚类法强调找出集合预报中的相似要素。

目前 ECMWF 使用管子法(Tubing clustering)对集合预报产品进行分类,管子法给出了天气预报的主要选择及其可能出现的天气现象,并提供了距离平均最远的极端天气图。这种方法首先对选定的区域内计算集合平均值,以集合平均为中心在给定的半径范围内所有成员为最基本的一组。然后以离集合平均最远的一个成员为准在相同大小半径的范围内的成员为另外一组。这一过程一直循环下去,每次分组数一般为 3~4 组,不能多于 5 组。这种方法的半径大小事先给定,且随季节不同而变化。这种聚类方法给出的预报精度取决于 Tubing 的分辨率,半

径越小,精度越高。管子法强调了集合中各元素之间的主要差别,它更注重集合的平均及其极值。

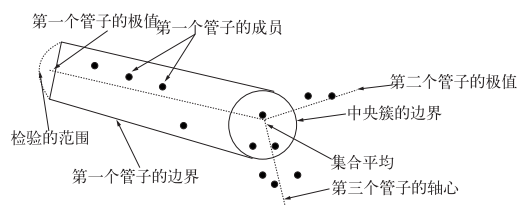


图 1 管子法示意图

中间的那一组的成员表示与集合均值最接近的一组

NCEP 采用距平相关系数分簇法对 500hPa 高度场进行分组,这种方法是将所有预报成员与历史平均值相减求出距平值,用距平值计算各成员之间的距平相关系数,找出所有成员中相似性最差(相关系数最小)的 2 个成员,这 2 个成员就给出了可能的预报范围,然后找出所有与这 2 个极端成员最相似(距平相关系数大于 0.6)的预报成员,确定为 2 类。对余下的预报成员,找出最相似的 2 个成员以及它们相似性较好的定义为另一类。这一过程循环下去,直到所有成员。这种方法强调天气形势的相似性。

法国气象局采用 Diday 提出的动力模糊法,初始划分时的重心由天气类型定义,划分用到的距离是位移和最大相关方法^[14-15],这种方法利用事先总结出的影响法国的天气类型,利用集合成员和重心之间的位移,将其分类。这一方法使预报员在集合预报系统和天气类型之间建立了直接联系,它为预报员提供了某种天气形势发生的概率,可以提高对重大天气过程的预报能力。但是由于实际天气有时是几个天气类型的叠加,因此,这种方法的局限性在于不能同时识别多个天气类型。

瑞典气象局使用神经元聚类法,这种方法基于神经网络原理。神经元是构成神经网络的最小单位,两个或两个以上的神经元就

能组成神经网络。神经网络的学习和识别取决于各神经元连接权系数的动态演化过程。同 Ward 聚类法,这种方法也是强调找出集合预报中的相似要素。

日本气象厅利用中央聚类法(Central clustering)进行聚类分析,就是我们通常所说的重心法,利用不同类之间重心间的距离作为标准进行划分,这种方法的缺点是重心不能代表单个样本的分布特点。

需要指出的是,ECMWF、NCEP 等主要中心的集合预报聚类分析方法在业务应用中也存在一定的局限性,由于集合预报系统在预报的中后期的发散度不够,导致了概率预报与观测到的实际发生的天气形势的频率不一致。

2 Ward 聚类法在中国业务集合预报系统中的应用

2.1 方案设计和业务化应用

中国业务集合预报系统是以中期预报模式 T213L31 为基础,采用繁殖增长方法生成扰动初值,由 15 个集合预报成员组成的新一代集合预报系统。该系统 120 小时以内预报时效的预报模式输出的时间间隔为 12 小时,120~240 小时预报时效的模式输出的时间间隔为 24 小时。

为了在我国业务集合预报系统中开发聚类产品,通过对 ECMWF 采用的 Ward 聚类法和 NCEP 业务实施的距平相关系数分簇法的实际应用对比试验,确定采用 Ward 聚类法来对我国业务集合预报产品进行分簇。2006 年 11 月,我国集合预报聚类产品开始投入业务使用,是对 T213 集合预报系统的 15 个成员利用 Ward 方法分为 3 类,并给出每一类的发生概率,每天 6:00(北京时)定时启动聚类分析作业,生成的聚类产品传输到中央气象台天气预报会商室的 MICAPS 服务器上,预报员可以通过 MICAPS 平台显示

和使用集合预报聚类产品。

2.2 应用个例分析

2006 年 11 月 7—9 日,我国淮河以北大部分地区出现了明显的雨雪、大风降温天气过程。本文以此次过程为例。选取这次过程集合预报起始时刻为 2006 年 11 月 5 日 12 时(世界时,下同)。图 2 给出了 11 月 5 日 12 时至 9 日 00 时的逐 12 小时 500hPa 环流分析场。可以看到,5 日 12 时亚洲中高纬度地区为两脊一槽型,东北高空冷涡伴随的长波槽控制着我国东部沿海地区,西西伯利亚至贝加尔湖为长波脊,极地冷涡盘踞在该长波脊以北的高纬地区,鄂霍茨克海地区亦为长波脊控制。从逐 12 小时的分析场环流形势演变来看,此次较强冷空气活动,主要是上游长波脊经向发展,脊轴由南北向逐渐转为东北西南向,脊前偏北气流分量加大,引导位于高纬地区的极地冷涡快速南下,8 日 12 时直抵东北地区形成东北冷涡,东北冷涡与其上下游长波脊构成小倒“Ω”型,因此,极地冷空气直袭我国偏东地区而且持续影响,淮河以北地区出现强冷空气活动过程,造成明显雨雪和大风降温。图 3 和图 4 分别为 48 小时和 84 小时集合预报的聚类产品(由于篇幅限制只给出两个时效的聚类分析产品),48 小时预报与相同时次的分析场对比,发生概率最高的图 3c 中环流形势配置和主要系统的形态与分析场最接近,只是主要影响系统高空冷涡的强度偏弱 40gpm;发生概率最低的图 3b 中不仅长波脊的走向因南北向与分析场不同,而且高空冷涡位置比实况更偏北、偏西,强度偏强 40gpm;发生概率次高的图 3a 中环流形势与实况的相似度介于图 3b 和图 3c 之间,但高空冷涡的强度与实况一致。对于图 4 中展示的 84 小时集合预报聚类产品,各类与分析场的对比分析结论与 48 小时预报的基本相同,发生概率最高的小倒“Ω”

与分析场最接近。综上所述, Ward 聚类法最高发生概率所对应的环流形势与分析场基本一致, 最低发生概率所对应的环流形势最偏离实际的环流形势, 说明了基于 Ward 聚类法的业务集合预报系统的聚类产品, 能够

有效地划分出最有可能发生的环流形势演变和调整, 给预报员提供有价值的预报参考信息, 同时也反映了采用 Ward 聚类分析方法开发的集合预报的聚类产品是合理的、可行的。

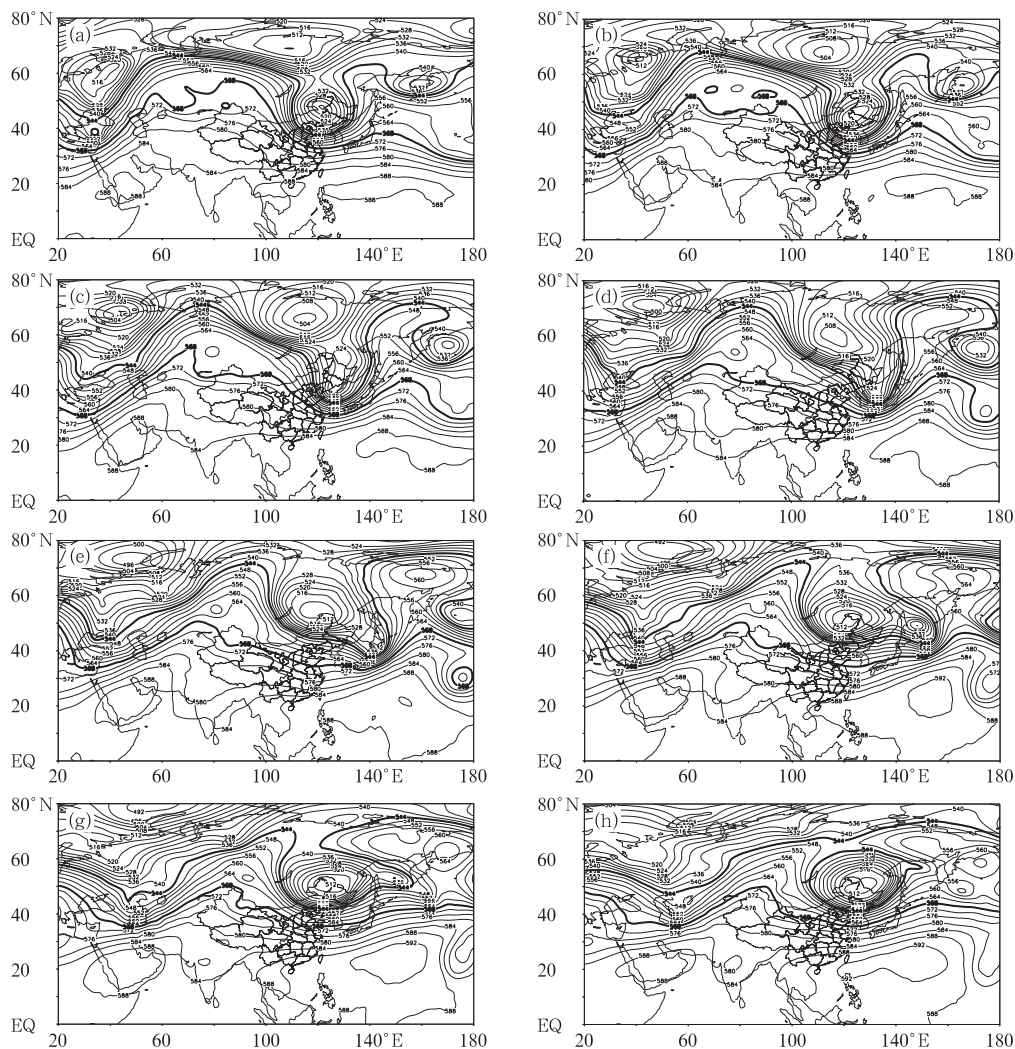


图 2 2006 年 11 月 5 日 12 时至 2006 年 11 月 9 日 00 时间隔 12 小时的 500hPa 环流形势分析场

另外, 本文也给出失败个例来说明集合预报聚类产品应用的局限性。2007 年 3 月 4—5 日, 受来自西西伯利亚强冷空气和强烈发展的江淮气旋北上共同影响, 东北地区出现了 1951 年有气象记录以来历史同期最强

的暴风雪天气过程。从分析场来看, 3 月 1 日 12 时, 500hPa 亚洲中高纬度地区为两脊一槽的强经向型, 乌拉尔山地区和鄂霍次克海地区分别为长波脊控制, 亚洲北部为宽广的低值区, 极涡中心位于 70°N 、 100°E 附近,

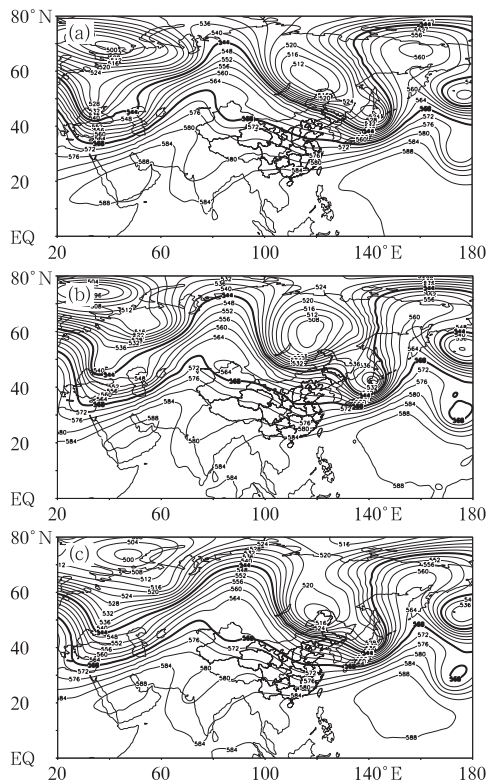


图3 2006年11月7日12时聚类分析结果
三组发生概率分别为40.0%、6.7%、53.3%

其伴随的长波主槽位于西西伯利亚地区。在随后的两天里,乌拉尔山长波脊东移,脊前偏北气流引导极地冷空气南下,高空冷涡及其伴随的长波槽东移发展,于4日08时抵达内蒙古东部,正是这个冷涡及其伴随的低槽东移活动为东北暴雪提供了冷空气源。同时,从青藏高原东移的高原槽与南支槽合并东移,不仅引导暖湿气流北上,而且槽前的正涡度输送为东北地区的降雪提供了动力条件。选取这次过程集合预报起始时刻为2007年2月28日12时(世界时),通过同时次集合预报的聚类产品与分析场比较发现,集合预报产品能够描述大尺度环流形势的演变特征,但对此过程预报的强度偏弱,冷空气移动的路径偏北、偏东。分析集合预报聚类产品预报失败的原因,可能由于(1)集合预报系

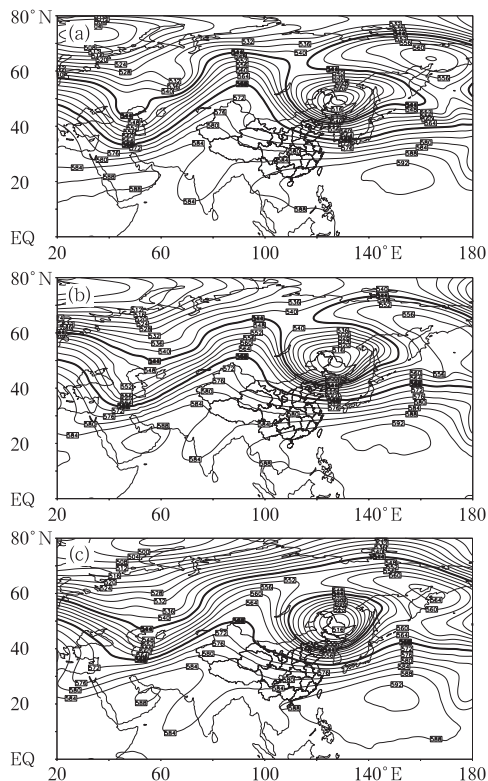


图4 2006年11月9日00时聚类分析结果
三组发生概率分别为26.7%、20.0%、53.3%

统本身不完美,天气系统发生发展过程复杂,以至于对于这次过程无法很准确描述;(2)Ward聚类分析方法强调的是集合预报成员间的相似性,目前给出聚类产品是组内集合成员的算术平均,这就有可能损失一些有用的信息,这在个别成员预报优于聚类分析产品的表现中得到验证。也说明,聚类分析产品仅仅是集合预报系统业务产品中一种表现方式,并不能取代所有成员罗列的邮票图产品。

3 结论和讨论

本文系统地介绍了集合预报技术先进国家业务集合预报系统中使用的集合预报聚类分析方法,同时,借鉴国外已有的技术,通过

实际应用对比试验,将 Ward 聚类法应用到我国业务集合预报系统中,并且进一步分析了集合预报聚类产品在重大天气过程中实际预报价值和应用效果。结果表明:

(1) 集合预报系统从原理上来讲是一个概率预报系统,涵盖未来可能发生的天气形势及发生的可能程度。开发集合预报系统的聚类分析产品的目的在于,希望通过对集合预报产品进行释用,找出最可能发生的天气形势,方便预报员使用。

(2) 国外集合预报技术先进的 NCEP 和 ECMWF 等气象部门对集合预报产品的解释应用技术的研发工作由来已久,比较深入,而在我国还处于起步和探索阶段,在集合数值预报代表数值预报未来发展方向的大趋势下,必须加大集合数值预报相关技术领域研究力度和投入,积极推动集合预报产品在我国业务部门中的广泛应用。

(3) 基于 Ward 聚类法的 T213 集合预报系统的聚类产品,能够有效划分出最有可能发生的环流形势演变和调整,给预报员提供有价值的预报信息。但这仅是一次重大天气过程的个例分析,全面科学评价 Ward 聚类分析方法在中国业务集合预报系统中的聚类分析质量和性能,需要利用多样本进行试验,通过天气学检验和统计检验相结合的方法来对分类结果进行检验,以验证这种方法的有效性。目前此方面的工作仍在进行中。

(4) 由于聚类方法和集合预报系统本身存在一定的局限性,在应用过程中也存在一些不足。首先,由于目前 T213 集合预报系统本身并不是完美的,再加上集合样本数的限制及实际天气变化的复杂性,使得集合预报的预报时效受到一定的限制。其次,因为天气形势的不同需要用不同的聚类分析方法,目前还没有理想的聚类方法可以满足业务天气预报对各类天气形势的需求,本文采用的方法只考虑不同成员间方差的不同,需要在今后的工作中开发出更能体现集合预报

科学内涵和具有天气学意义的聚类分析方法。

参考文献

- [1] 陈静. 中尺度暴雨短期集合预报研究[D]. 北京:中国气象科学研究院,2003.
- [2] Epstein E S. Stochastic Dynamic Prediction[J]. *Tellus*, 1969, 21(6):739-759.
- [3] Leith C E. Theoretical skill of Monte Carlo forecasts[J]. *Mon. Wea. Rev.*, 1974, 102:409-418.
- [4] Toth Z. and Kalnay E. Ensemble forecasting at NMC: The generation of perturbation[J]. *Bull. Amer. Meteor. Soc.*, 1993, 74:2317-2330.
- [5] Zoltan Torh and Eugenia Kalnay. Ensembl forecasting at NMC: the generation of perturbations [J]. *Bull Amer Meteor Soc*, 1993, 74(12):2317-2329.
- [6] 杜钧. 集合预报的现状和前景[J]. *应用气象学报*, 2002,13(1):16-28.
- [7] 皇甫雪官. 国家气象中心集合数值预报检验评价[J]. *应用气象学报*, 2002,13(1):29-36.
- [8] 陈静,陈德辉,颜宏. 集合数值预报发展与研究进展[J]. *应用气象学报*, 2002,13(4):497-507.
- [9] 田华. 国家气象中心集合预报系统概况[J]. *新疆气象*, 2004,27(5):1-6.
- [10] 李泽椿,陈德辉. 国家气象中心集合数值预报业务系统的发展及应用[J]. *应用气象学报*, 2002, 13(1):1-5.
- [11] Ward J H. Hierarchical grouping to optimize an objective function[J]. *J. mer. tat. ssoc.* 1963, 58: 236-244.
- [12] Atger F. Tubing: an alternative to clustering for ensemble prediction classification[J]. *Wea. Forecasting*, 1999, 114(5):741-757.
- [13] 施能. 气象科研与预报中的多元分析方法[M]. 北京:气象出版社,2002:228.
- [14] 杨学胜. 位移和最大相关距离在 ECMWF 集合预报产品中的应用[J]. *应用气象学报*, 2002, 13(1):37-45.
- [15] 杨学胜. 根据天气类型划分欧洲中期天气预报中心的集合预报产品[J]. *气象学报*, 2001, 59(2):173-182.
- [16] 李小泉,刘金达,汪迎辉. 集合预报及其在中期天气预报中的应用[J]. *气象*, 1997, 23(8):3-9.