

刘家峻,张立凤,关吉平,等. 集合预报产品在台风麦莎预报中的应用[J]. 气象,2010,36(5):21-31.

集合预报产品在台风麦莎预报中的应用^{*}

刘家峻¹ 张立凤² 关吉平² 李荔珊¹ 费增坪¹

1 中国人民解放军 61741 部队

2 解放军理工大学气象学院, 南京 211101

提 要: 为了提高集合预报效果,利用 T106L19 全球谱模式和增长模繁殖法对台风麦莎制作了 13 个成员的集合预报,并对其结果进行集合预报产品解释应用研究。首先在离散距离分簇法中,引入方差分析的方法确定分簇基础,然后对集合预报的形势场进行了分簇,相对于以往采用 96 小时样本为分簇基础进行分簇的方法,分簇效果提高明显,分簇结果能给出台风移动的几种可能路径及其概率;其次对要素随时间的演变利用系统聚类法进行了分类,根据分簇平均图可以确定几个可能的台风登陆地点,制作了这些特殊点要素分类烟羽图,通过分析要素随时间的演变特征,提高对台风登陆地点和时间的预报精度;最后制作了盒须图,通过盒须图中数据组主体以外的数据点,确认出值得注意的一些特殊集合成员的预报结果,减少了小概率事件的漏报率。研究结果表明,将方差分析引入离散距离分簇法,要素烟羽聚类法以及盒须图的应用有利于提高和改善集合预报效果,集合预报产品解释应用效果得到改进。

关键词: 集合预报, 台风, 分簇, 烟羽聚类图, 盒须图

Application of Ensemble Prediction Products to the Forecast of Typhoon Masta

LIU Jiajun¹ ZHANG Lifeng² GUAN Jiping² LI Lishan¹ FEI Zengping¹

1 No. 61741 Troop of PLA

2 Institute of Meteorology, PLA University of Science and Technology, Nanjing 211101

Abstract: To improve the effect of ensemble prediction, a 13-member ensemble forecast of Masta was generated by using T106L19 model and breeding growth method(BGM), and the research of the application of ensemble prediction products from the results was carried out. Firstly, the pattern fields of ensemble forecast were clustered through variance analysis according to discrete distance between members. Through this method, the pattern fields are grouped into three different clusters, especially the several possible tracks and the proportions of every cluster member are provided. The 96 h forecast results were used as the basis of clustering in the previous method of discrete distance clustering. In comparison with the previous method, the results of variance analysis are better. Secondly, the evolvement of variables is categorized by using the method of systemic classifying and several possible typhoon landing positions can be concluded from clustering averaged plots. Based on these special points, the plumes plot with systemic classifying of variables can be used to analyze the sea level pressure of the special points. Through discussing the evolvement characteristics of these points, the prediction results of the landing time and position of typhoon are more similar to observations than the cluster average. In addition, box-whisker plot has been presented. Several special ensemble members have been identified through the analysis of data beyond the

^{*} 国家重点基础研究项目“我国南方致洪暴雨监测与预测的理论和方法研究(2004CB418304)”资助,解放军理工大学气象学院博士启动资金资助

2008 年 12 月 01 日收稿; 2010 年 1 月 29 日收修定稿

第一作者:刘家峻,主要从事暴雨中尺度动力学诊断分析与数值模拟. Email:liujiujun81@sohu.com

main distribution of data. The information provided by the special points from box-whisker plot can reduce the missing rate of severe weather event. The research results indicate that the effects of prediction and the application of ensemble prediction products were improved by using the method of the clustering basis determined by using variance analysis, systemic classifying and box-whisker plot.

Key words: ensemble prediction, typhoon, cluster, plums with systemic classifying, box-whisker plot

引言

由于大气是一个对初值敏感的、高度非线性的动力系统,所以像决定性数值预报方法那样,仅用众多可能的预报状态的一种来描述大气未来的状态是不够科学的,更合理的方法是以概率的形式来认识大气预报状态,而对所有可能的预报状态及这些预报状态的概率分布进行预报,这种预报方法就是集合预报。集合预报为解决单一确定性预报存在不确定性的问题提供了一条新的途径。通过对集合预报产品解释应用方法的研究,能提高对极端性天气的预报能力和对危险性天气的防御能力,为预报员的预报决策提供有效的、实用的预报结果,这些成果将具有应用价值。目前发达国家都开展了集合预报业务,且多种集合预报产品已应用到实际业务中,其中以欧洲中心的集合预报产品最为丰富,包括离散度图、要素概率图、邮票图、概率烟羽图等^[1-4],在数值预报产品中集合预报产品占 60%。我国对集合预报产品的研究也越来越受到重视,中国气象局的中期集合预报也实现了业务化,但对集合预报的解释应用工作远远不够,产品数量还不够丰富,解释应用方法还不够完善。目前相关方面的学术论文也很少,皇甫雪官^[5]对集合预报结果进行了 Talagrand 概率分布、离散度、BS 评分、命中率以及空报率的统计检验。毛恒青等^[6]采用面条图、温度变化趋势图等对集合预报产品进行了解释应用。本文将针对几种集合预报产品应用作进一步的研究,并将其应用到“麦莎”台风的预报中,以期提高对极端性天气的预报水平,使集合预报结果更加客观^[7-10]。

1 台风麦莎和试验方案简介

2005 年 7 月 31 日至 8 月 9 日,0509 号台风麦莎影响了我国东部沿海和华东地区,使我国出现大范围暴雨和大风天气,此次台风是继 1997 年 11 号台风后,近 8 年来对我国影响最严重的台风之一,给

人民的生命财产造成很大损失^[11]。

台风麦莎于 2005 年 7 月 31 日在菲律宾以东洋面上生成,在东南气流引导下向西北方向移动,8 月 3 日 02 时(北京时间,下同)加强为台风,8 月 6 日 03 时 40 分在浙江玉环登陆,登陆时中心附近最大风速 $45 \text{ m} \cdot \text{s}^{-1}$,然后向西北方向移动,8 月 6 日,江苏省受到“麦莎”前锋的影响,普降暴雨,8 月 6 日 22 时 15 分后,“麦莎”影响安徽,8 月 7—8 日,“麦莎”席卷山东,8 月 7 日晚,天津出现 2005 年首次风暴潮,8 日出现一次大范围的强降水天气。9 日 07 时 10 分在辽宁省大连市龙王塘镇再次登陆。台风登陆后开始一直沿西北方向移动,在 7 日 08 时,当其移动到 31.2°N 、 118°E 附近时,又转为向东北向移动一直经过山东半岛到达渤海湾^[12-14](见图 1)。

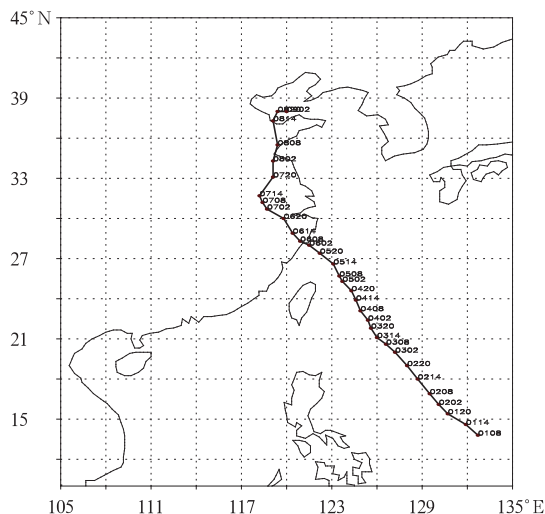


图 1 2005 年 8 月 1 日 08 时到 9 日 02 时
台风麦莎实况路径(间隔 6 小时)

Fig. 1 Best track of typhoon Masta from 08:00
BT 1 August to 02:00 BT 9 August 2005
(time interval: 6 h)

采用 T106L19 全球谱模式和增长模繁殖法对本次台风过程制作了集合预报,繁殖循环初始扰动取为 T106L19 模式 24 小时预报均方根误差乘在 $[-1, 1]$ 之间均匀分布的随机数;繁殖模采用成对繁殖的方法生成,即繁殖模定义为加扰预报与减扰预

报之差的 $1/2$;繁殖循环周期同客观分析周期取为 6 个小时;总时间繁殖为 3 天;前人研究表明^[15-16],在集合预报中,采用 8 个集合成员求平均已能完成集合平均对预报精确度改善量的主要部分,增加集合成员数有利于提高集合预报效果,由于计算条件限制,在此集合成员取为 13 个,其中一个控制预报,6 个加扰预报和 6 个减扰预报。预报时效为 10 天,前 5 天为每 6 小时输出一次结果,后 5 天为每 12 小时输出一次结果,共有 31 次结果。

2 集合预报产品的应用

2.1 形势场分簇

根据集合预报各成员的演变趋势,可以把所有成员预报的结果分成几簇,使每一簇的变化趋势十分相似,每一簇的平均场就是形势场分簇图。这是有效提取集合预报产品有用信息的方法之一。该产品可以为预报员提供几种类型的形势场演变趋势,并给出每种类型可能出现的概率大小。欧洲中期天气预报中心(ECMWF)采用差值均方根分簇法进行分簇,该方法将各成员之间 500 hPa 高度场的差值均方根作为衡量它们之间距离大小的标准,分类时选定欧洲区域,是以 5~7 天的预报场作为分簇的基础^[6]。我国现在采用的分簇法是离散距离分簇法,该方法以各成员对集合平均的离散程度大小进行分类,采用 $30^{\circ}\sim 50^{\circ}\text{N}$ 、 $110^{\circ}\sim 130^{\circ}\text{E}$ 范围内 500 hPa 高度场 96 小时预报资料作为分簇基础^[17]。具体做法如下^[6]:

设每个预报成员在选定区域内的格点高度值分别为 H_{ij} ($i=1,2,\dots,N$, $j=1,2,\dots,M$),其中 i 为预报成员序号, j 为选定区域格点的序号。则区域内各点上高度的平均值为:

$$\bar{H}_j = \sum_{i=1}^N H_{ij} / N \quad (1)$$

各成员在各点上的距平值为:

$$\Delta H_{ij} = H_{ij} - \bar{H}_j \quad (2)$$

区域内各成员的平均距平值为:

$$\overline{\Delta H}_i = \sum_{j=1}^M \Delta H_{ij} / M \quad (3)$$

因此,在该区域内所有成员的离散距离定义为:

$$L = \max(\overline{\Delta H}_i) - \min(\overline{\Delta H}_i)。$$

根据离散距离大小,将预报成员分为如下 3 簇:

第一簇: $\overline{\Delta H}_i \leqslant -L/4$; 第二簇: $-L/4 < \overline{\Delta H}_i \leqslant +L/$

4; 第三簇: $+L/4 < \overline{\Delta H}_i$ 。

2.1.1 方差分析在离散距离分簇法中的应用

采用离散距离分簇法,分簇基础的确定是分簇成功与否的关键,目前没有一个统一的确定方法,前面已经提到离散距离分簇法中以往取第 96 小时的预报场作为分簇的基础。为了找出一种更合理确定分簇基础的方法,我们将方差分析引入离散距离分簇法中。方差分析是通过试验数据的分析来辨别矛盾主次的一种数学方法,方差分析中的 F 统计量反映了簇间差异与簇内差异之比,若 F 值越大,则说明同簇的样本数值的离散程度小,而不同簇的样本之间数值离散程度大,则可以说明分簇越显著。计算 F 统计量公式为^[18]:

$$F = \frac{\bar{S}_1}{\bar{S}_2} \quad (4)$$

其中: $\bar{S}_1 = \frac{S_1}{a-1}$, $\bar{S}_2 = \frac{S_2}{\sum_{i=1}^a b_i - a}$, 而 $S_1 = Q - P$,

$S_2 = R - Q$, P 、 Q 和 R 的计算方法如下:

$$P = \frac{1}{\sum_{i=1}^a b_i} \left(\sum_{i=1}^a \sum_{j=1}^{b_i} y_{ij} \right)^2 \quad (5)$$

$$Q = \sum_{i=1}^a \frac{1}{b_i} \left(\sum_{j=1}^{b_i} y_{ij} \right)^2 \quad (6)$$

$$R = \sum_{i=1}^a \sum_{j=1}^{b_i} y_{ij}^2 \quad (7)$$

其中, i 表示簇号,范围从 1 到 a ; b_i 表示第 i 簇内的样本数; j 表示簇内样本号,范围从 1 到 b_i ; y_{ij} 表示第 i 簇第 j 号样本。 S_1 的自由度为 $a-1$; S_2 的自由度为 $\sum_{i=1}^a b_i - a$ 。

以 8 月 2 日及 8 月 4 日起报的 850 hPa 高度场的 10 天预报为例,分簇基础的计算区域为 $15^{\circ}\sim 50^{\circ}\text{N}$ 、 $90^{\circ}\sim 150^{\circ}\text{E}$,考虑到预报时效为中期时段以及 T106 模式的预报效果,将 3~8 天模式输出的 18 个时次的集合预报结果作为分簇的基础,按以上公式计算 F 值。图 2 为 F 值曲线图。分析图 2 可以看出,对 3~8 天的预报结果,8 月 2 日起报的 F 最大值出现在第 66 小时(即模式第 12 个时次的输出结果);8 月 4 日起报的 F 最大值出现在第 72 小时(第 13 个时次)。也就是说对于这两天的预报以 66 小时和 72 小时作为分簇基础进行分簇时,可以使簇内成员相似性最大,簇间成员的相似性最小。

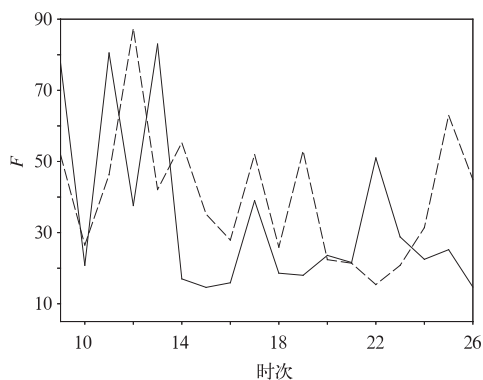


图 2 2005 年 8 月 2 日和 4 日起报模式
第 3 到第 8 天 F 值随时间的演变
Fig. 2 The evolvement of F -value from
the third day to the eighth day with time
(the initial time: 00 UTC 2 August
and 00 UTC 4 August 2005)

2.1.2 分簇结果分析

表 1 给出的是以第 66 小时预报场作为分簇基础,对 8 月 2 日起报的 10 天预报的 850 hPa 高度场分簇的结果。

表 1 2005 年 8 月 2 日起报的 850 hPa 高度场分簇结果
Table 1 The cluster result of 850 hPa geopotential height
(the initial time: 00 UTC 2 August 2005)

簇序号	簇样本	占成员总数的百分比
第一簇	1-, 2-, 5+	23.1%
第二簇	控制预报, 3-, 4-, 6-, 4+, 6+	46.2%
第三簇	5-, 1+, 2+, 3+	30.8%

注: $x-$ 表示第 x 个减扰预报, $x+$ 表示第 x 个加扰预报, x 取 1~6

8 月 2 日起报是提前 4 天对台风登陆进行预报,模式控制预报及集合平均预报的结果是台风中心都不会登陆我国,而在各成员预报中,第 4、5 个减扰预报以及第 3 个加扰预报都报出了比控制预报好的结果(图略)。从表 1 可见,第三簇成功地将第 5 个减扰预报和第 3 个加扰预报分离出来。为了分析台风登陆过程,下面只给出了台风登陆前后即预报的第 90 小时到第 240 小时三簇 850 hPa 高度场簇平均随时间演变图(图 3),其中,图 3a、3b、3c 分别为第一、二、三簇的簇平均时间演变图。在簇平均图中,第三簇报出了台风在预报的第 132 小时登陆,比

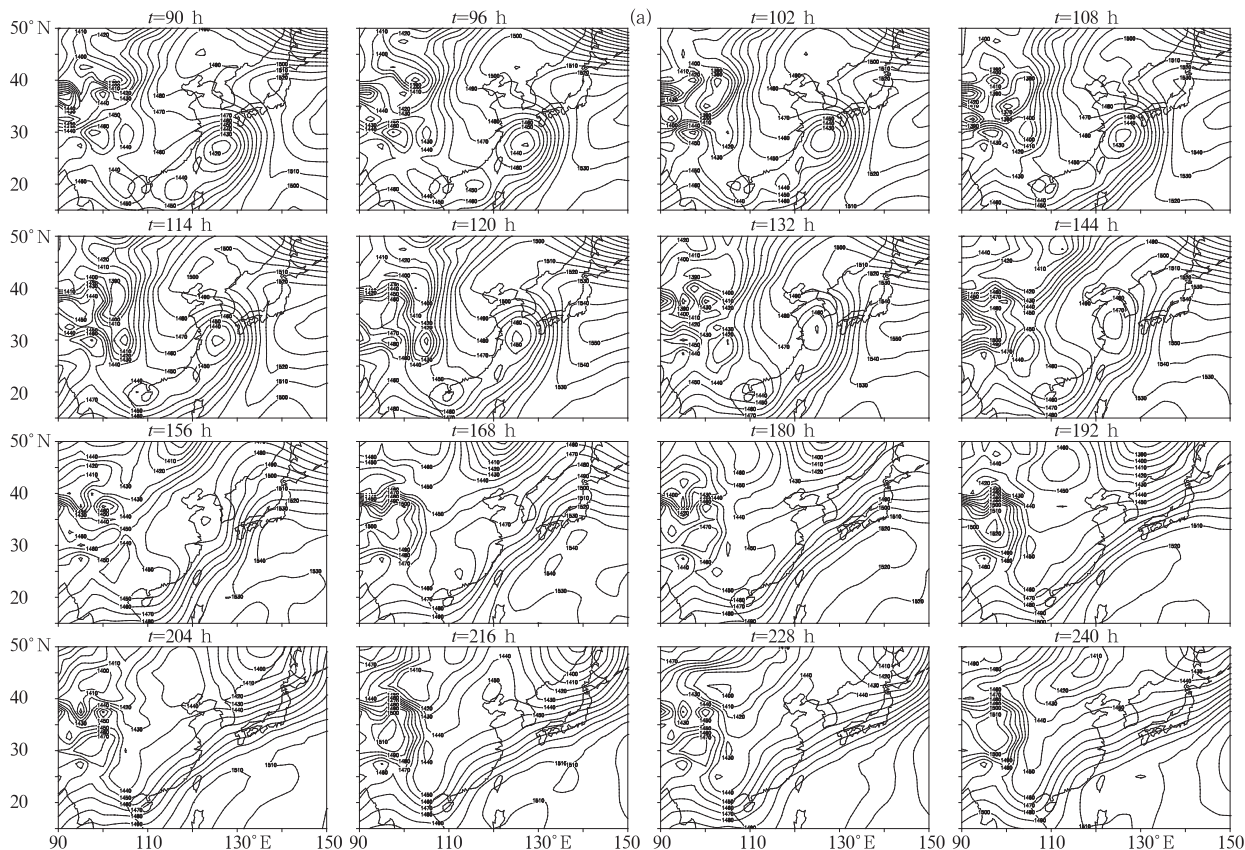
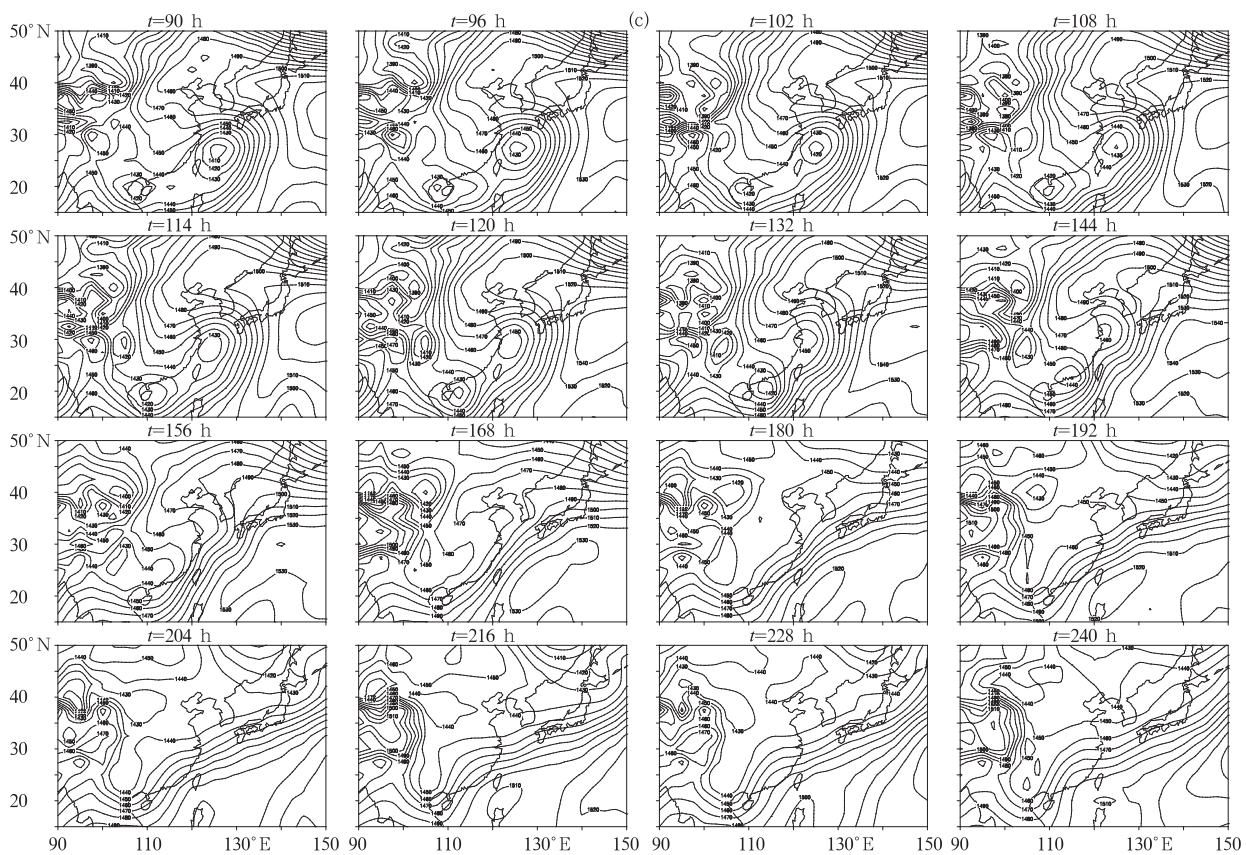
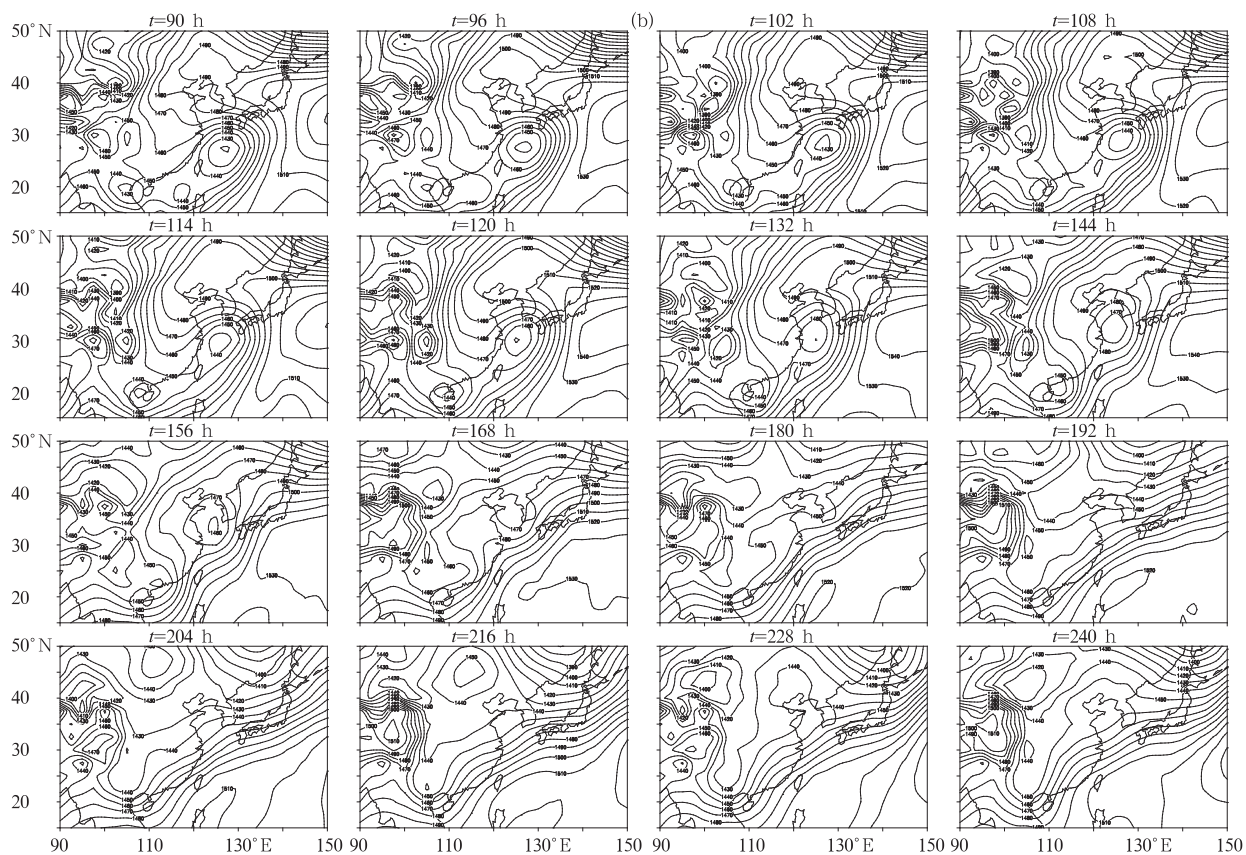


图 3 2005 年 8 月 2 日起报 850 hPa 高度场簇平均时间演变图

(a) 第一簇; (b) 第二簇; (c) 第三簇

Fig. 3 The evolvement of 850 hPa geopotential height cluster average with time
the initial time: 00 UTC 2 August. (a) the first cluster, (b) the second cluster, (c) the third cluster



续图 3

实况要晚,且在图中台风登陆后没有闭合环流,这可能与模式分辨率及集合预报输出结果时间密度不够有关。第三簇对于台风登陆后向西北方向移动,其后又转向东北方向移动的实况过程没有报出,但实际上第三个加扰预报报出了这一转向的移动路径,簇平均漏报台风转向可能是由于簇成员平均后弱信息被平滑掉的结果。由此可见分簇平均图给出了集合平均预报所不能给出的信息,但一些特殊成员预报也不可忽视。此外特别要注意的是,第三簇所占比重虽然不是最大的,但第三簇包括了集合成员预报较好的成员样本的 66.7%,整个簇比重达到 30% 以上,所以这一簇的预报结果应给予重视。

表 2 给出的是以第 72 小时预报场作为分簇基础,对 8 月 4 日起报的 10 天预报的 850 hPa 高度场分簇结果。

8 月 4 日 08 时起报模式,对台风登陆(6 日 03 时)是提前近 2 天的预报。控制预报和集合平均预

报的结果相近,其结果是台风中心都不会到达我国内陆地区。图 4 给出了 30~120 小时三簇 850 hPa 高度场的簇平均随时间演变图。图 4a、4b、4c 分别对应表 2 中的第一、二、三簇的簇平均时间演变图。其中只有第一簇报出了台风的登陆及其后的转向移动,但与实况相比,登陆时间推迟了 12 个小时左右,登陆地点偏北。其余两簇都没有报出台风到达我国内陆,其预报路线为在海面沿西北向移至江淮地区,然后沿海岸线北移到达山东半岛,后两者的差别在于台风中心距离大陆远近。在 13 个成员预报中第

表 2 2005 年 8 月 4 日起报的 850 hPa 高度场分簇结果

Table 2 The clustering result of 850 hPa geopotential height (the initial time: 00 UTC 4 August 2005)

簇序号	簇样本	占成员总数的百分比
第一簇	1-, 2-, 5+	23.1%
第二簇	控制预报, 3-, 6-, 4+	30.8%
第三簇	4-, 5-, 1+, 2+, 3+, 6+	46.2%

注: $x-$ 表示第 x 个减扰预报, $x+$ 表示第 x 个加扰预报, x 取 1~6

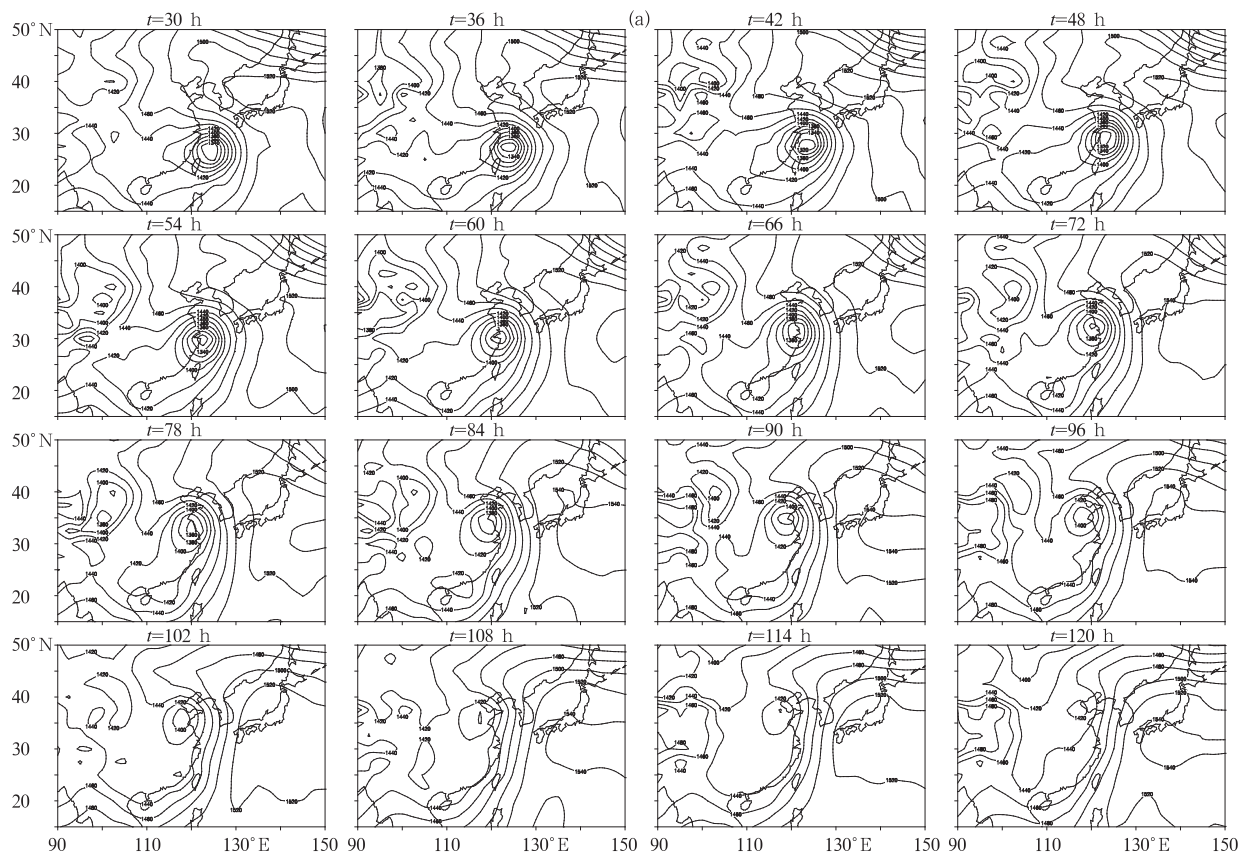
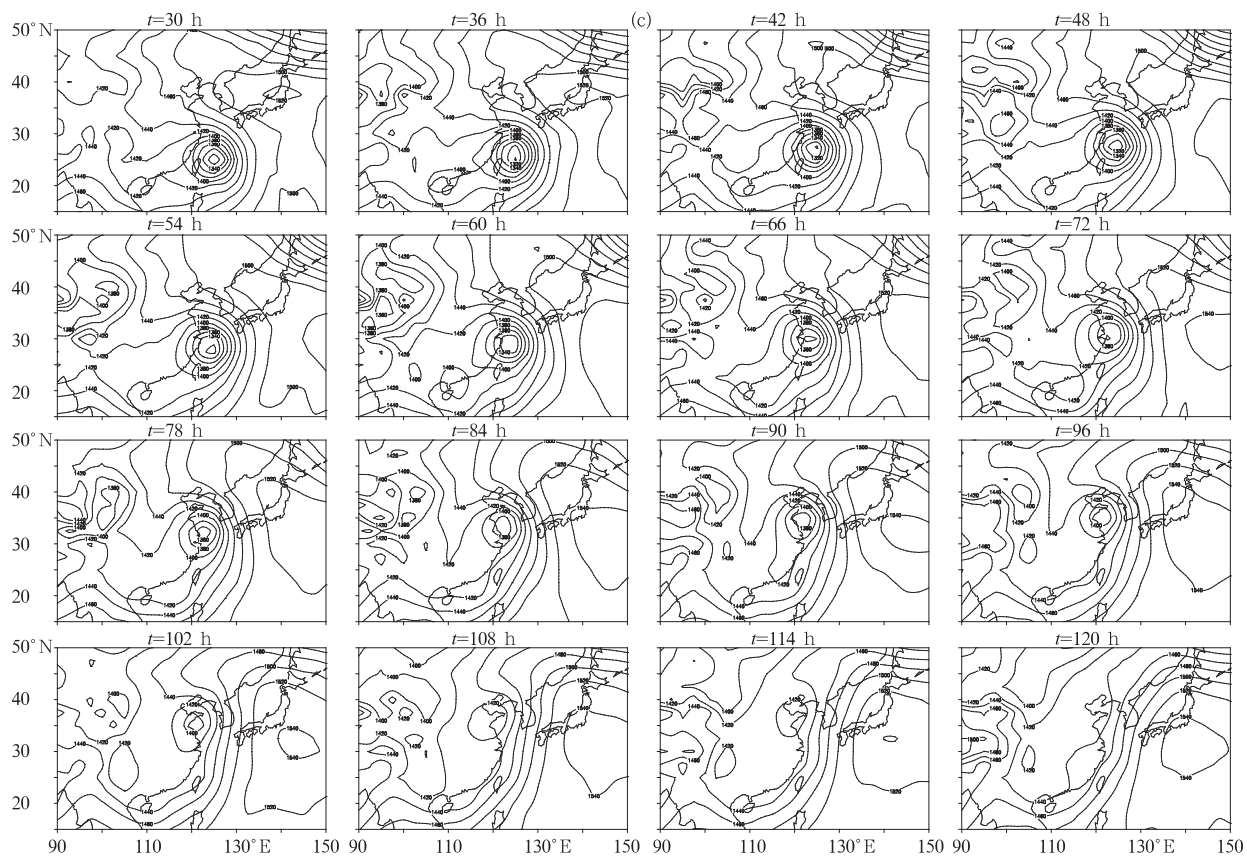
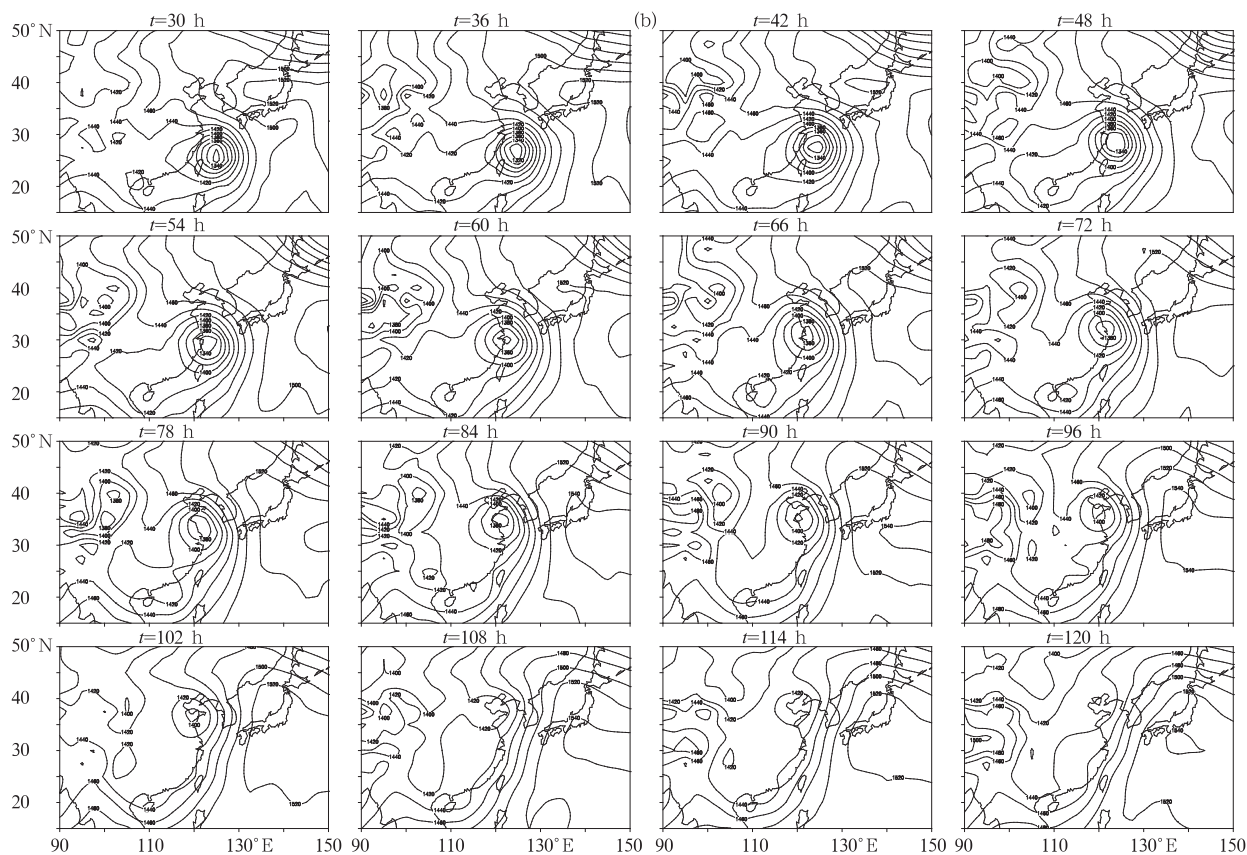


图 4 2005 年 8 月 4 日起报 850 hPa 高度场簇平均时间演变图

(a)第一簇; (b)第二簇; (c)第三簇

Fig. 4 The evolvement of 850 hPa geopotential height cluster average with time

the initial time: 00 UTC 4 August 2005. (a) the first cluster, (b) the second cluster, (c) the third cluster



续图 4

5 个加扰预报、第 1、2、4、6 个减扰预报都成功地预报出了台风登陆以及登陆后先向西北方向移动然后转向东北方向移动的情况。这 5 个成员中有 3 个成员包括在第一簇中。

2.1.3 与不采用方差分析的结果对比

如果采用以往常用的 96 小时预报结果作为分簇基础,8 月 2 日和 4 日起报的分簇结果见表 3 和表 4。

表 3 2005 年 8 月 2 日起报以 96 小时为分簇基础的 850 hPa 高度场分簇结果
Table 3 The clustering result of 850 hPa geopotential height through the data of 96 h (the initial time: 00 UTC 2 August 2005)

簇序号	簇样本	占成员总数的百分比
第一簇	1-, 4-	15.4%
第二簇	控制预报, 2-, 3-, 5-, 4+, 5+, 6+	61.8%
第三簇	6-, 1+, 3+	23.1%

注: $x-$ 表示第 x 个减扰预报, $x+$ 表示第 x 个加扰预报, x 取 1~6

表 4 2005 年 8 月 4 日起报以 96 小时为分簇基础 850 hPa 高度场分簇结果
Table 4 The clustering result of 850 hPa geopotential height through the data of 96 h (the initial time: 00UTC 4 August 2005)

簇序号	簇样本	占成员总数的百分比
第一簇	1-	8%
第二簇	控制预报, 2-, 3-, 4-, 5-, 1+, 2+, 3+, 4+, 5+, 6+	84.6%
第三簇	6-	8%

注: $x-$ 表示第 x 个减扰预报, $x+$ 表示第 x 个加扰预报, x 取 1~6

8 月 2 日起报的三簇比重分别为 15.4%、61.8%、23.1%, 这三簇的簇平均都没有报出台风的登陆。从表中可以看到, 预报效果较好的第 4 个、第 5 个减扰及第 3 个加扰的成员预报分别被分在 3 个簇中。

8 月 4 日起报的三簇的比重为 8%、84.6%、8%。只有第一簇报出了台风的登陆及转向, 但比起应用方差分析法得到的三簇中第一簇的比重 23.7%, 可信度小了很多。由此可见, 用 F 检验确定分簇基础是十分必要的, 这将大大提高分簇的效果。

2.2 气压场烟羽聚类图

概率烟羽图是为了表示在某一固定地点某个气象要素的集合预报离散度在整个预报期限内的变化情况。具体做法是选取某一关心地点, 统计计算集合预报各成员对此点上气象要素的概率分布, 然后将所有预报时效的要素概率绘制成等值线图, 由于

其形状很像烟雾和羽毛, 故叫概率烟羽图^[6]。一般而言概率烟羽分布范围较大时, 可信度较差; 当烟羽出现分叉状态时, 大气处于不确定状态^[19]。本文制作的是气压场概率烟羽图, 其中气压间隔取 1 hPa。

在前面 8 月 4 日起报的分簇个例中, 第一簇簇平均预报结果是在第 54 小时(图 4a 的第 5 张小图)台风大致在 (30°N、122°E) 登陆。由于分簇法仅从大体上提供了几种形势场的可能演变趋势, 台风登陆地点的确定还很粗略。为了缩小这一误差, 在 30°N 附近的海岸线上, 分别取 A(31°N、122.0°E), B(30°N、120.75°E), C(29°N、121.75°E), D(28°N、120.75°E), E(27°N、120.5°E) 五点, 通过制作和分析这些点上的气压烟羽聚类图, 来进一步确定台风登陆的地点。具体作法是: 首先利用双线性插值法将集合预报的海平面气压插值到这 5 个点, 然后将每个点上集合平均预报的海平面气压随时间的演变绘在一张图中(见图 5)。

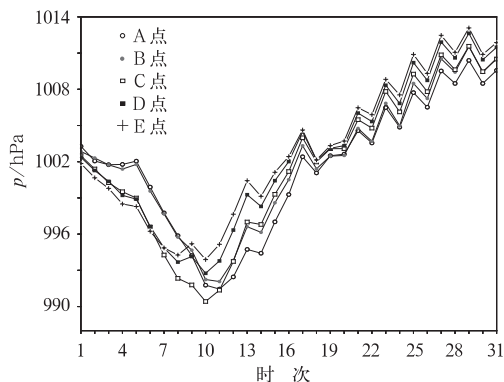


图 5 2005 年 8 月 4 日起报海平面气压随时间演变
Fig. 5 The evolvement of sea level pressure with time (the initial time: 00 UTC 4 August 2005)

由于台风是热带地区的暖性低压涡旋, 中心的气压值很低, 当台风环流靠近某一地区时, 该地气压开始缓慢下降, 当接近台风中心时, 气压出现陡然下降, 当达到台风中心时气压达到最低值, 台风经过后, 气压便迅速回升^[20]。利用台风过境气压的这种变化规律, 分析图 5 上五点海平面气压的演变, C 点的海平面气压随时间下降的幅度最为明显, 在 8 月 6 日 14 时达到了气压最低值, 由此推测台风有可能在 C 点登陆, 实际上 C 点与台风实际登陆地点玉环 (28.14°N、121.23°E) 十分接近, 仅偏北不到 100 km。气压达到最低值的时间与台风的实际登陆时间 8 月 6 日 03 时相差大约 11 个小时, 产生误差的一种可能原因是模式输出时间密度不够。

选择 C 点作海平面气压概率烟羽图(如图 6)用

以分析在该点各成员预报的离散度。其中,“+”线代表的是控制预报,“。”线代表的是集合平均预报。阴影区域代表海平面气压的概率。阴影的大值区出现在初始时刻、第18小时(图中对应第4个时次)及第168小时(第25个时次)到216小时(第29个时次),概率值大于30%,要素预报可信度较高。另外,在第48小时(第9个时次)到第66小时(第12个时次)阴影区对应的概率值也达到了20%以上。在第24小时到第78小时要素分布较分散,出现多

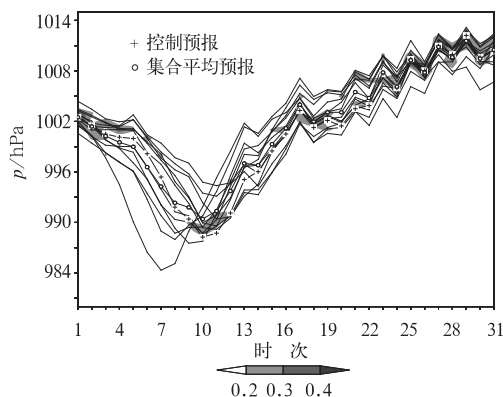


图6 2005年8月4日起报C点
海平面气压概率烟羽图

Fig. 6 The probability plume plot of sea level pressure on the point C
(the initial time: 00 UTC 4 August 2005)

分叉现象,即大气处于不确定状态,对这一时段的预报需多加注意。而这正处于台风登陆前后。

为进一步掌握在分叉处海平面气压的离散状况,我们采用系统聚类法中的最短距离法对各成员海平面气压随时间演变趋势进行分类。

系统聚类法是目前国内外常用的一种方法,其基本思想是先将每个样品各自看成一类, n 个样品即有 n 类,选择距离最小的两类并成一个新类,计算新类和其他类的距离,再将距离最近的两类合并,距离计算公式如下^[21-22]:

$$x_{ijk} = x_{ik} - x_{jk} \quad (8)$$

$$E_{ij} = \frac{1}{(m_2 - m_1 + 1)} \sum_{K=m_1}^{m_2} x_{ijk} \quad (9)$$

其中, i, j 代表样本号,取值范围为 $1 \sim 13$; K 代表离散度大的时次,这里取第5~14个时次,即 $m_1 = 5$, $m_2 = 14$; E_{ij} 代表了 i 样本对 j 样本各时次之间的总平均差值。

$$S_{ij} = \frac{1}{(m_2 - m_1 + 1)} \sum_{K=m_1}^{m_2} |x_{ijk} - E_{ij}| \quad (10)$$

S_{ij} 能反映出两个样本中各个时次之间的差值 x_{ijk} 对 E_{ij} 的离散程度,能较好地反映两样本间的形相似程度。

图7给出了用系统聚类法对C点海平面气压进

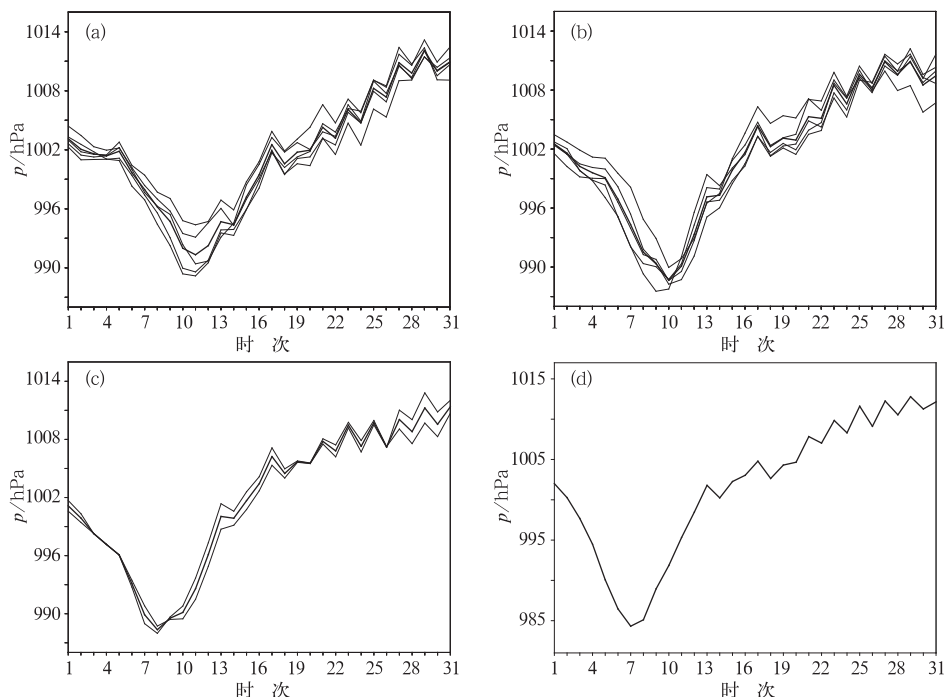


图7 2005年8月4日起报C点海平面气压概率烟羽聚类图
(a) 第一类; (b) 第二类; (c) 第三类; (d) 第四类

Fig. 7 The probability plume plot with systemic classifying of sea level pressure on the point C
the initial time: 00 UTC 4 August 2005. (a) the first category, (b) the second category, (c) the third category, (d) the fourth category

行分类的结果,其中图 7a、7b、7c、7d 分别代表第一、二、三、四类,粗线条代表类平均,四类中类成员占总集合成员数的百分比分别为 38.5%、38.5%、15.4%、7.7%。

通过分析几类图中海平面气压最低点出现的时间,可以预报台风登陆的时间。在图 7 中,第一类最低点出现在第 60 小时(图中对应第 11 个时次),第二类最低点出现在第 54 小时(第 10 个时次),第三类最低点出现在第 42 小时(第 8 个时次),第四类最低点出现在第 36 小时(第 7 个时次)。尽管第三类所占比重不是最大的,但第三类最低点出现时间最接近实际台风登陆时间,这反映了在业务预报中,这几类情况都不能忽视。

2.3 盒须图(box-whisker plot)

盒须图能提供数据变化范围和极端值的信息。方盒表示中间 50% 数据的位置,须向盒外两侧延伸至 1.5 倍半极差范围内最远的数据点^[20]。盒须图具体的制作方法为:从一个四分位数到另一个四分位数画一盒子,并在中位数处画一横线穿过框。然后从框形的每一端画一竖直虚线,竖直虚线止于这样一个区间的极值上,该区间为(下四分位数 $-1.5 \times$ 四分位数间距,上四位数 $+1.5 \times$ 四分位数间距)^[23]。

盒子中的横线表示的中位数是表明数据位置的一个很好的指标,它在各种情况下都能保持稳定,也就是说中位数是一个稳健估计。使用稳健估计的优势在于,在实际中大多数物理数据组并不是正态分布的,因此样本均值不再是最佳估计量。另一方面,长数据组中不可避免地包含一些远离中心的数据,这时样本均值对于这些数据来说并不具有很好的代表性^[24-25]。对于集合预报来说,难免有一些远离中心的数据,稳健估计可以使要素预报的估计值不受远离中心数据的影响。两个四分位数与中位数的差值可以给出数据偏斜度的某些信息。盒子长度越长,说明数据分布越分散,预报的不确定性越大。除此之外,根据盒子各部分长短的不同,可以判断数据集中的值域。利用两个极值、两个四分位数和一个中位数这样 5 个数值就能很好地概括集合要素预报^[24]。

图 8 给出了 8 月 4 日起报 C 点海平面气压演变的盒须图,其中小圈代表(下四分位数 $-1.5 \times$ 四分位数间距,上四位数 $+1.5 \times$ 四分位数间距)区间之外且在(下四分位数 $-3 \times$ 四分位数间距,上四位数

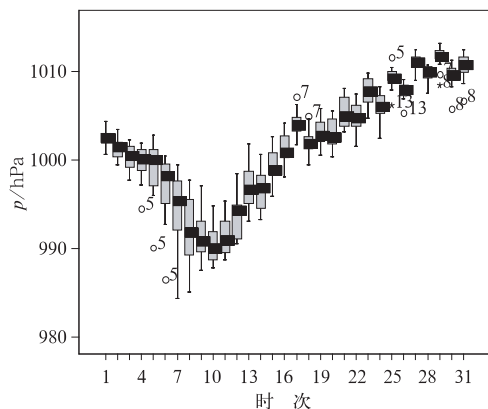


图 8 2005 年 8 月 4 日起报 C 点
海平面气压盒须示意图

Fig. 8 The whisker-box plot of sea level
pressure on the point C
(the initial time: 00 UTC 4 August 2005)

$+3 \times$ 四分位数间距)区间之外的值,星号代表(下四分位数 $-3 \times$ 四分位数间距,上四位数 $+3 \times$ 四分位数间距)区间之外的值,图中数字表示成员编号。这样保证在正态分布中有 95% 的数据均位于盒形图(盒子和两端的须)内。因此,在盒形图外的点能详细地表示集合成员中的一些特殊成员。这样找到的特殊成员的预报结果可能是正确的,也可能误差较大。检验出特殊成员后,不能只作简单的删除。很多情况下,这些成员往往包含比其他成员更多的信息量,必须根据实际情况作仔细分析。

分析图 8 可见,在第 36 小时(图中对应第 7 个时次)到第 48 小时(第 9 个时次)盒子长度及盒子外的须伸展的范围都很大,可由此推测在这一时段预报处于较强的不确定状态,需重点分析,这正对应台风登陆前期。随着盒子长度由长变短再变长,对应了台风的登陆前期、登陆和登陆后的移动。其中 5、7、8、13(对应第 4、6 个减扰预报以及第 1、6 个加扰预报)这几个成员在显著性检验范围之外,它们很可能是一些与主体数据相差较远的特殊成员,也可能是单一确定性预报所不能提供的重要信息。盒须图中离散度较大时段内第 5 个成员(即第 4 个减扰预报)自第 18 小时(第 4 时次)至第 30 小时(第 6 个时次)一直在显著性检验之外,由前面分析已知,第 4 个减扰预报报出了台风的登陆和转向,所以对于这些点也不能忽视。

3 小 结

本文利用 T106L19 全球谱模式和增长模繁殖

法,制作了13个成员的中期集合预报,针对“麦莎”台风登陆过程,对集合预报产品进行了解释应用,由于篇幅限制,并没有对影响台风路径的关键天气系统进行预报性能分析,这一内容将另外撰文研究。主要结论如下:

(1) 为提高分簇效果,将方差分析引入离散距离分簇法中,可使分簇基础更加客观和合理,从对850 hPa高度场的集合预报进行分簇的结果看,这种方法可以将不同的预报结果分离出来,特别是对预报效果比较好的成员能分在一簇中,并给出各簇的比重,这与以往采用96小时预报场作为分簇基础的分簇结果相比,效果得到明显改善。

(2) 对台风麦莎的预报,850 hPa高度场可分为三簇,其中一簇的簇平均结果与实况接近。

(3) 根据簇平均图中得到台风可能登陆地点,利用不同点的海平面气压烟羽图,通过分析海平面气压随时间的演变特征,可以进一步订正和确定台风的登陆地点和登陆时间。台风麦莎的试验结果表明,对登陆地点和登陆时间的预报结果都比簇平均图的预报更接近实况。

(4) 对要素烟羽图中多分叉时段利用系统聚类法进行分类,得到预报点海平面气压的烟羽聚类图,反映了几种可能的台风登陆时间。台风麦莎的试验结果被分成四类,同时给出了每一类占集合预报的比重。

(5) 盒须图不仅给出了以往集合预报常用的信息,还给出了数据组主体以外的特殊数据点,这些点可能是大气状态的极端情况,反映了单一确定性预报所不能提供的重要信息,利用这些点可以减少小概率事件的漏报率。

参考文献

- [1] 蔡其发,张立风,张铭. 中期数值天气预报的集合预报实验[J]. 气候与环境研究,1999,4(4):365-373.
- [2] 关吉平,黄泓,张立风. 集合预报中初始扰动生成方法的探讨[J]. 解放军理工大学学报,2003,4(2):87-90.
- [3] 杜钧. 集合预报的现状和前景[J]. 应用气象学报,2002,13(1):16-28.
- [4] 杨学胜. 业务集合预报系统的现状及展望[J]. 气象,2001,27(6):3-9.
- [5] 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.
- [6] 毛恒青,陈谊,陈德辉. 基于神威中期集合数值预报系统的产品开发[J]. 应用气象学报,2002,13(1):47-51.
- [7] 李俊,纪飞,齐琳琳,等. 集合数值天气预报的研究进展[J]. 气象,2005,31(2):1-7.
- [8] 严明良,缪启龙,沈树勤. 基于超级集合思想的数值预报产品变权集成方法探讨[J]. 气象,2009,35(6):19-25.
- [9] 黄燕燕,万齐林,袁金南,等. 基于BDA扰动法的台风路径集合预报试验研究[J]. 热带气象学报,2006,22(1):49-54.
- [10] 陈静,陈德辉. 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.
- [11] 肖风劲. 降水偏多气温接近常年台风麦莎造成损失严重[J]. 气象,2005,31(11):94-95.
- [12] 高栓柱,孟智勇,杨贵名. 台风麦莎渤海转向的可预报性研究[J]. 气象,2009,35(2):8-14.
- [13] 蒋名淑,施丹平,商兆堂,等. 台风麦莎的天气过程总结[J]. 成都信息工程学院学报,2006,21(4):571-576.
- [14] 顾华. 华北中南部出现强降水“麦莎”、“珊瑚”登陆我国[J]. 气象,2005,31(11):89-93.
- [15] Epstein E S. Stochastic dynamic prediction[M]. Tellus, 1969, 21: 739-759.
- [16] Houtemamer P L, Derome J. Methods for ensemble prediction[J]. Mon Wea Rev, 1995,123:2181-2196.
- [17] 田华. 国家气象中心中期集合预报系统概况[J]. 新疆气象,2004,27(5):1-3.
- [18] 黑龙江泰来县气象科. 用方差分析选择预报因子[J]. 气象,1975,1(3):18-20.
- [19] 金荣花,李维京,孙照渤. 集合预报产品释用方法的研究[J]. 南京气象学院学报,2002,25(4):442-448.
- [20] 杨信杰,钱家声,陆胜元,等. 天气学原理[M]. 解放军理工大学气象学院,1987,276-308.
- [21] 陈谋国. 统计天气学原理和方法[M]. 解放军理工大学气象学院,1987,182-214.
- [22] 李开乐. 相似离度及其使用技术[J]. 气象学报,1986,44(2):174-182.
- [23] 王键,靳奉祥. 盒须图和相关模型法数据质量控制[J]. 山东科技大学学报,2002,21(2):55-58.
- [24] Allan H Murphy, Richard W Katz. 大气科学中的概率统计和决策. 北京:气象出版社,1988:6-22.
- [25] 高栓柱,张守峰,钱传海,等. 基于位置误差的分布制作热带气旋路径袭击概率预报[J]. 气象,2009,35(9):38-43.