

华东台风降水预报的一种集成^{*}

贡九鼎 钮学新 李慧武

(浙江省气象局, 杭州 310002)

提 要

如何改进预报集成是当前业务预报上急需解决的问题, 本文给台风降水定义了重解。根据重解对众多预报结果作自动筛选, 最终由人机对话对少数近邻的预报结果作出集成。该集成能远离平均态, 适用于大的台风降水之预报。

关键词: 集成; 台风; 重解; 近邻。

一、引 言

当前要素的业务预报基本上是采用一些数值预报产品的统计预报。最终预报结论是各种预报方法结果的集成。由于参加集成的个体多, 而且有的个体之间差别较大, 加之以往的集成结果又是一个期望值, 故预报不出有特色的结果。由此, 在台风降水预报上, 预报结果总是在平均态附近振动。这种集成结果尤其满足不了业务上对较大降水预报的需要。可见, 要改进集成效果, 就要减少参加集成的个体数目, 并减小个体之间的差异。

另一方面已经积累了大量的历史资料, 以至要找到 2—3 个降水相近邻的台风已不困难。事件之后, 总能找到 2—3 个历史台风, 用它们的平均值作预报可以很好地逼近当前态的实况。这说明, 大量的历史资料已提供了不少有关当前态的解的信息^[1]。

这里的信息指历史台风的总降水量及其分布, 还有登陆地段。它们自然也是该历史台风的解。这样, 每一个符合起报条件的历史台风都是一个已知解。如何由少数几个相近邻的已知解集成为当前态的解即是本文之目的。

设一个变量对应一个解, 或者说解 x_i 由变量 a_i 输出。如果将一个变量看成一种预报方法, 那么下面的论述原则上适合于多种预报方法的集成。

譬如采用相似预报方法, 对当前态台风的一组变量得到一组已知解, 已知解中一般只有少数几个解和当前态的未知解相近邻。由于这少数几个近邻解在事件之前是未知的, 故我们面临的是一个“一多对应”的因果关系。如果在变量组中有一个如此重要的变量, 以至解每次可由它大致确定, 那自然能在欧几里德空间中找到唯一近邻, 即唯一解。这不啻是一个尽管粗糙但却实用的“一一对应”关系。但事与愿违, 在台风降水预报上经多年实践整理出来的若干变量都能以不同概率成为这种主要变量, 从而还是“一多

1992年4月14日收到, 11月9日收到再改稿

^{*} 中国气象局台风基金资助课题

对应'的因果关系。

把“一对应”关系转化成“一少对应”关系,而且少数结果之间的差别要尽可能小,这是本文的指导思想。只有这样,才能集成出一个有特色的解,该解有远离平均态的可能。

实际工作中还有一种情形,即对当前态的解起决定作用的变量只是众多变量中的少数几个变量。而且这些少数变量随当前态的不同而不同。解往往由少数几个变量决定的情形,似乎反映了哈肯奴役(slaving)原理——状态变量之间的合作和竞争最终导致只有少数慢变量起支配系统的作用^[2]。

本文用筛选手段,由变量在筛选过程中“幸存”的时间长短定义慢快,与哈肯的定义有所不同。

二、解集结构和变量选取

设解由总降水量和登陆地段构成。本文采用的等雨量线是:50、100、200、300、400 (mm)。 a_0 、 a_1 、 a_2 、 a_3 、 a_4 分别是上面等雨量线所覆盖的面积。则每个台风的总雨量可表为 $r = 50a_0 + 100a_1 + 200a_2 + 300a_3 + 400a_4$ 。再将 r 分成 5 个等级,从小到大以 1、2、3、4、5 表之。

台风的降水量和登陆地段关系密切,尤其是较大的台风降水。就华东地区而言,在浙江海门到福建厦门之间登陆的台风降水强度较强,范围较大。其主要原因是台风倒槽发展得最完善,影响范围也最广。在上述地段以北登陆的台风,其倒槽位置偏北。在上述地段以南登陆的台风则与西风带系统相距较远,台风倒槽的发展和影响范围受到抑制。

根据以上基本事实,将登陆地段作数字化处理,见表 1。

表 1 登陆地段代码

登陆 地段	崇明 象山	象山 海门	海门 温州	温州 福州	福州 泉州	泉州 漳州	漳州 汕头	汕头 香港	
代码	2	3	4	5	5	5	4	3	2

数码仅表示了该地段的最大概率降水量级,而且均未考虑登陆地段和雨型的关系(表中有 3 个不同的登陆地段均用数码 5 表之)。这样实际上将路径雨量化了。对 124°E 以东北上、珠江口以西登陆(包括进入南海)的台风均以 1 记之。对 124°E 以西北上的台风以 2 记之。对西进点达浙闽近海 150km 以内北上或消失的台风,均作其为最西端纬度的登陆台风处理。

令 r_k 与 t_k 表示解 x_k 的总降水量和路径。由于降水和路径均已数字化,而且路径的数字化是根据降水大小而来的;因此 r_k 与 t_k 是可加的,即有 $x_k = r_k + t_k$ 。

两个解 x_i 与 x_j 之间的距离是 $|x_i - x_j| = |r_i - r_j| + |t_i - t_j|$ 。 $|x_i - x_j| = 0$ 表示 x_i 与 x_j 为重解。

1. 解集结构

考虑一个变量输出一个解的简单情形。 p 个变量值 $a_1, a_2, \dots, a_p$ 对应当前态 A 的一个由 p 个解元素构成的解集 $X = (x_1, x_2, \dots, x_p)$ 。对任意解元素 $x_k \in X$ 定义距离 $dx_k = \min |x_k - x_j|$, $i = 1, 2, \dots, p$; 且 $i \neq k$ 。则 X 就是一个距离空间。

可用单位圆直观地表示解集 X 的结构。上面已将解 x_k 定义成一个自然数, 故 dx_k 也是一个自然数, 它当然可被映射成一个角度 θ_k , 即 $\psi(dx_k) = e^{i\theta_k}$ 。若以单位圆中的半径代表解, 则解与解之间的关系就是半径之间的夹角关系。 $\min$ 表示取两个夹角之中小的一个。

如果 $dx_1 = dx_2 = \dots = dx_p$, 即 $\theta_1 = \theta_2 = \dots = \theta_p$, 此时 X 为对称结构。对称结构的解集中每个解出现的几率一样, 为 $1/p$ 。

如果 x_k 为重解, 即 $dx_k = \theta_k = 0$, 则单位圆在该方向上产生对称破缺^[3], 解 x_k 在解集 X 中的出现几率立即陡增到 $2/p$ 。三重解出现的几率为 $3/p$ 。

重解的出现, 表明多个坐标变量输出同一个解。其结果使 X 中不同解的数目减少, 出现了从“一多对应”向“一少对应”的转化。

再作一直观说明。设解集 X 只含 6 个解, 则解集的结构分三种情形。见图 1。

图 1a 中, 各解距离(夹角)相等, 形成了对原点的中心对称。当前态的变量值和 6 个解构成了“1-6 对应”的因果关系。实际上 6 个解中和当前态未来实况相近邻者只是其中的少数几个解。由于解集结构是对称的, 作集成时只能平均。故集成结果必然是一个没有特色可言的平均结果。

图 1b 中, 6 个解叠在一起, 是 6 重解。原中心对称的图形在此破缺。这是一种理想的“一一对应”的因果关系。集成结果当然有特色。

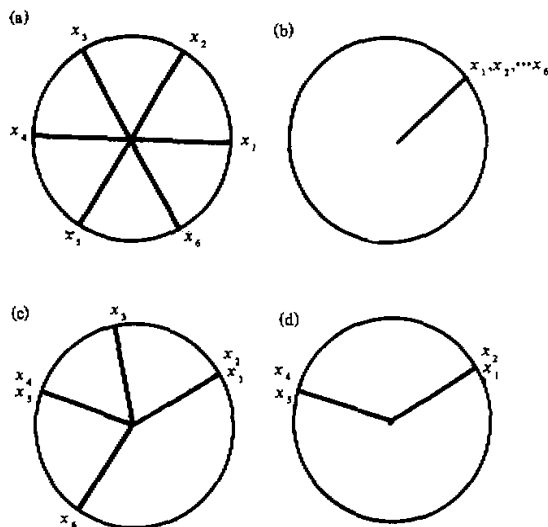


图 1 解集结构示意图

图 1c 中有二个重解, 是实际上常遇见的情形。这是“1-4 对应”的关系。从概率上讲, 当前态的解宜取二个重解之一。因此, 图 1c 又反映了“1-2 对应”的因果关系。输出解 x_3 、 x_6 的坐标变量 a_3 、 a_6 自然成为无关的快变量。将快变量 a_3 、 a_6 消去, 再作一轮输出运算, 就得到图 1d 的解结构。需指出, 图 1d 上的解与图 1c 上对应变量所输出的解不一定相同。

图 1d 中两个重解之间的距离还是大的, 若作集成, 仍然得不到有特色的结果。不过选择范围已大大缩小, 从“1-6 对应”转化到“1-2 对应”。

不论对 x_1 与 x_2 作集成, 抑或对 x_4 与 x_5 作集成, 其结果都具有特色。至于集成结果是否和当前态实况相近邻, 那是另一个问题。

2. 坐标变量

(1) 台风特征量: 中央台 12—24 小时路径综合预报, 12—24 小时台风中心强度预报。

(2) 欧洲中心 48 和 72 小时 500 hPa 高度场预报值及变高值。将这两种值在下面三个区中作切比雪夫展开: 西风带槽脊活动区 (30—50°N, 100—140°E), 副热带高压南北轴向区 (10—40°N, 130—150°E), 热带气旋影响区 (15—40°N, 110—140°E)。从中挑出 6 个切比雪夫系数作变量。

(3) 下述指标站风矢: 青岛、南京、上海、杭州、衡水、温州、福州、厦门、南昌、赣州、汕头、银川、延安、西安、太原、北京、郑州、宜昌、济南 (以上取 500 hPa 实况风); 黄山、天目山、括苍山、七仙山、射阳、石埭、宫古岛、冲绳、奄美大岛、济州岛 (以上取地面风)。共计 29 个站的风矢, 依据浙闽两省预报员的经验作 6 种组合, 得到 6 个变量。

总共 14 个变量是从更多的变量中选取的。选取原则是在每一个历史解集里, 泉州以南和温州以北登陆的台风不能并存。

三、求解过程

求解过程就是变量的“幸存”过程。过程中, 将输出最孤立解 (其距离最大) 的变量作为快变量逐一剔除。

1. 自动筛选求重解

求当前态 A 的第一个解集 X^m 。

设 DE 为辗转筛选算子^[4], 将其作用在 A 和 p 个变量上得解集 X^0 ,

$$DE(A/p) = (x_1^0, x_2^0, \dots, x_p^0) = X^0.$$

求 X^0 中最孤立的解元素 x_i^0 , 它应当满足

$$dx_i^0 = \max_k \min_l |x_k^0 - x_l^0|. \quad (k, l = 1, 2, \dots, p; \quad k \neq l)$$

将输出 x_i^0 的坐标变量 a_i 视为快变量消去, 注意, 消去的是变量 a_i 而不是解 x_i^0 。

对 A 在 $p-1$ 个变量上重复上述过程。

$$DE(A/p-1) = (x_1^1, x_2^1, \dots, x_{p-1}^1) = X^1,$$

又可消去一个快变量; ... 直到 $DE(A/p-m) = (x_1^m, x_2^m, \dots, x_{p-m}^m) = X^m$ 。此时

$$dx_l^m = \max_k \min_l |x_k^m - x_l^m| = 0. \quad (k, l = 1, 2, \dots, p-m; \quad k \neq l) \quad (1)$$

(1) 式表示解集 X^m 中只剩下重解, 相当于完成从图 1c 到图 1d 的过程。上角标 m 记下了消去的快变量数目, 同时也表示了慢变量“幸存”时间的长短。输出解 $x_1^m, x_2^m, \dots, x_{p-m}^m$ 的变量 $a_1, a_2, \dots, a_{p-m}$ 就是慢变量。

要十分注意, X^m 中的 a_i 与前几步解集中的 a_i 不一定相同。至于 x_l^m 与 x_l^0 一般不相同就更不用说了。

求 A 的第二个解集 Y^n 。

实际工作中, 针对某一个具体问题, 往往有一个变量虽非决定性的, 但相对说来是重要的。譬如台风路径对台风降水的作用。然而也存在一些原因有使重要的变量在最初几步就被当作快变量消去的可能。为此, 可将程序稍作改动, 以保证其不被消去。在这种条件下, 重复上述过程。

$$\begin{aligned} DE(A/p) &= (y_1^0, y_2^0, \dots, y_p^0) = Y^0 \\ DE(A/p-1) &= (y_1^1, y_2^1, \dots, y_{p-1}^1) = Y^1 \\ &\vdots \\ DE(A/p-n) &= (y_1^n, y_2^n, \dots, y_{p-n}^n) = Y^n \end{aligned}$$

此时

$$dy_l^n = \max_k \min_l |x_k^n - x_l^n| = 0, \quad (k, l = 1, 2, \dots, p-n, \quad k \neq l) \quad (2)$$

Y^n 中只剩下重解。重解输出的流程见图 2。

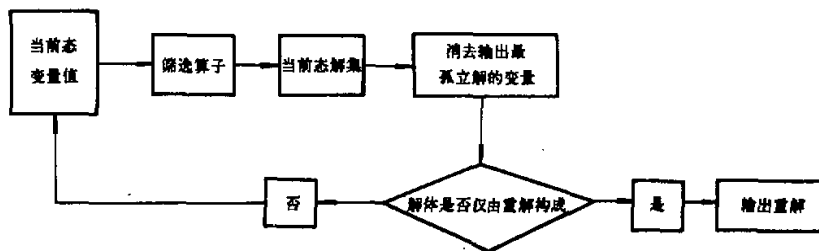


图 2 重解输出流程

2. 人机对话作决策

如果变量和解之间确有相当理想的关系, 应该有 $X^m \approx Y^n$ 。事实上只是 X^m 中的某些解与 Y^n 中的某些解相近邻。为了得到当前态的一个收敛的解, 应在 X^m 与 Y^n 中挑出 2—3 个最近邻的解作集成, 则集成结果将会具有特色。

这里的“近邻”不能靠本文定义的重解去衡量,因为重解的定义太粗糙。譬如登陆地段数码只是根据总降水量的概率而定的,没有考虑降水分布——即雨型(见表1)。由于温州以北的解和泉州以南的解不会出现在同一个解集之中(否则变量要作调整),故尽管数码一样也不会引起混淆。但是在温州—泉州之间登陆的台风,只要雨量级别相同就是重解。这种重解里的雨型显然会相差很大。即使用其它方法能将温州—福鼎,福鼎—福州,福州—泉州三个地段加以区别,也远不能描述实际上千姿百态的等雨量线走向。譬如登陆前后的路径对降水的影响有时比登陆地段更重要。对上述复杂的情形,宜采用人机对话方式。将 X^m , Y^n 的解从资料库中调出,直接从全程路径和雨量实况图中去找“最近邻”者。

整个求解分三个步骤:

第一步,由已知解求变量。1956—1989年符合起报条件^[4]的历史台风共150个,即有150个已知解。依次将每一个台风作当前态求解,得到150个解集。要求每个解集里泉州以南(包括进入南海)登陆的台风与温州以北(包括海上转向北上)登陆的台风不能并存。根据这一原则,从众多变量中选取了如第二节中的14个变量。14个变量对应14个解元素,组成一个台风的解集,这是“1-14对应”的因果关系。

第二步,求当前态的重解。在求重解的过程中逐一消去快变量,达到“1-少对应”的因果关系。

由于重解的定义相当宽,解之间的差别尚不小;尤其在温州—泉州之间登陆的台风在雨型分布和集中降水落区上还有较大的区别,故“人机对话”是不可缺少的最后一个步骤。

四、效果分析

1. 操作步骤

以9015台风为例。9015一经踏上起报线即可读出14个变量值。那么 $DE(A/p) = (x_1^0, x_2^0, \dots, x_p^0)$ 对应的解集是 $DE(9015/14) = (6126_{(4)}^0, 6205_{(2)}^0, 6214_{(5)}^0, 6513_{(4)}^0, 6615_{(5)}^0, 5622_{(4)}^0, 5612_{(3)}^0, 5905_{(4)}^0, 6404_{(1)}^0, 8707_{(4)}^0, 8909_{(3)}^0, 8506_{(4)}^0, 7209_{(4)}^0, 7413_{(3)}^0) = X^0$ 。以上14个解均在泉州到象山之间登陆,各种降水级别都有,但集中在4级(见解的右下标)。

根据第二节“1. 解集结构”中的定义,求每个解的距离。将输出最孤立的解6205的变量 a_2 (48小时500 hPa高度场预报值的一个切比雪夫系数)作为快变量消去。返回再做 $DE(9015/13) = (x_1^1, x_2^1, \dots, x_{13}^1) = X^1$, 参见图2。注意 $x_i^1 = x_i^0$ 不一定成立。经消去10个快变量后得到9015的第一个解集 $X^{10} = (6513, 6615, 5622, 5905)$ 。

6513, 5622均在福建平潭登陆,降水量级都是4,是一组重解。6615, 5905分别在福建霞浦和连江登陆,降水量级都是4,也是重解。故解集 X^{10} 中的解元素满足式(1)。分别输出该4个解的变量是: x_1 ——西风带槽脊区500 hPa高度场的一个切比雪夫系数, x_2 ——福州、温州、衢县、赣州、南昌、杭州、上海风矢的SE分量之和, x_3 ——12—24小时中央台的台风路径综合预报, x_4 ——黄山、天目山、括苍山、七仙

山东风风速之和。该 4 个变量就是最后“幸存”的慢变量。

同理, 9015 的第二个解集 $Y^9 = (6404, 5612, 8909, 6404, 6404)$ 。6404 是一个三重解, 它在浙江近海的大陈岛消失。5612, 8909 分别在浙江象山、海门登陆, 降水量级都是 3, 应视为重解。故 Y^9 中的解元素满足式 (2)。 Y^9 中的 5 个慢变量有 4 个与 X^{10} 一样, 外加一个 500hPa 副热带高压南北轴向变高场的切比雪夫系数(y_2)。

解集 X^{10} 与 Y^9 之间距离较近的解是 6615、5905、5612、8909。从资料库中将这 4 个台风的路径、雨量资料调出; 立见 8909 与其它 3 个有明显差异。故取 6615、5905、5612 的平均值作 9015 的预报集成, 见图 3。图 4 是 9015 的实况。

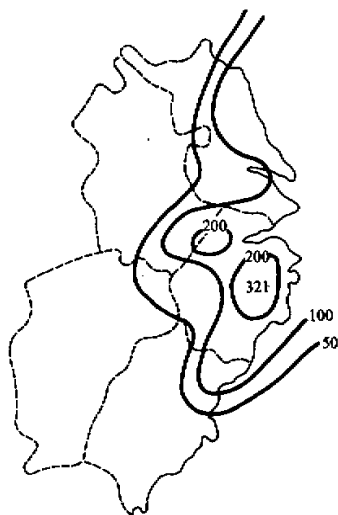


图 3 9015 预报集成

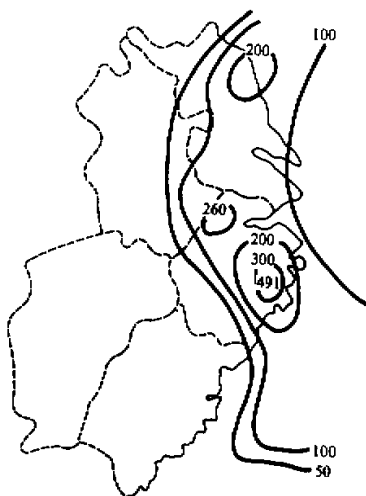


图 4 9015 雨量实况

要注意, 输出 5612 的变量 y_2 在 Y^9 中是慢变量, 而在 X^{10} 中却是快变量。

2. 与欧几里德距离和回归方程作比较

若变量不经第三节的“幸存”过程, 可直接得出在变量 x_2 、 x_4 、 y_2 上与 9015 有最近欧几里德距离的解: 8712, 8807, 6404。变量一样, 但变量输出的解不一样。欧几里德距离的 3 个解差异大, 平均后得不到有特色的结果, 见图 5。

因为降水少的样本多, 用回归方程报不出 4—5 级降水。回归方程中的变量是不随当前态改变的。在最小二乘原理下得出的回归系数也不随当前态改变。只靠当前态的变量值形成不了巨涨落。当然回归本身也是一种集成, 只是该种集成结果往往在均值附近, 无特色可言, 见图 6。

3. 拟合和预报

若预报集成的降水量级和实况一致, 且两者最大雨量点相距在 100km 之内算一次成功的预报, 则 150 个历史台风的拟合率为 77%。拟合失误的 35 个台风中只有 9 个是

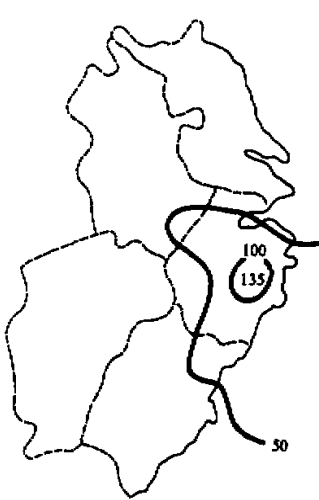


图 5 8712、8807、6404 雨量平均

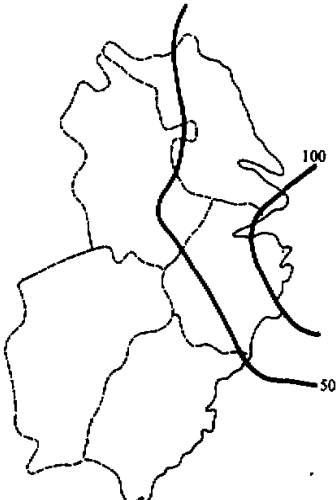


图 6 回归方程结果

在华东登陆的。值得注意的是，这 35 个台风中有 24 个给出了不能满足式 (1) 或 (2) 的信息。这一点在实际工作中是有用的。它提醒预报员，对该次台风宜采用其它途径作预报。

以 X 表示 1990—1991 年业务预报结果。 X^e 、 X^r 分别表示欧几里德距离和回归方程对同一个台风的预报结果。见表 2。

表 2 实况与三种预报结果比较

量级 \ 编号	9007	9012	9015	9018	9021	9107	9109	9110	9117	9120
实况	1	4	4	4	1	1	1	2	1	1
X	1	3	4	4	1	1	无解	2	无解	1
X^e	2	3	3	2	2	3	2	3	1	1
X^r	2	3	3	2	2	2	1	2	1	2

因为符合起报条件的样本中，降水小的居多数（占 3/4 强），故可将数码 1、2 看成众数。数码 3 作平均态看。从表 2 知， X^e 、 X^r 的结果集中在 1、2、3 三个数码上； X^r 更是集中在 1、2 二个数码上。唯有 X 有离开众数和平均态的现象。

五、小结与讨论

对当前态的一组变量值 $\{a_i\}$ ，经历“一多对应”、“一少对应”、“少中求近”等一系列过程，最终得到当前态的一个较为收敛的解集。从文献[4]知，消去的是变量 a_i 而不是解 x_i ； x_i 的数目在“幸存”过程中不变。 a_i 之间的相互作用如文献[4]中所述。 x_i 之间也相

互作用。该相互作用首先体现在从全体已知解中有哪几个 x_i 被变量 a_i 输出而成为当前态的可能解元素。再者, 每消去一个变量 a_i , 全体解 $\{x_i\}$ 又在更深层次上作一番竞争后, 有选择地被输出。因为消去一个变量后, 就更能突出其余变量的相互作用以及变量 a_i 与解 x_i 之间的相互作用。另外, 决定每个当前态的解的(慢)变量都不尽相同, 解之间就会形成较大的涨落。故本文的集成方法对较大降水的预报能力胜过一般统计方法, 但仍有下述问题。

(1) 并非一定是最佳集成

9015 台风预报集成的最大雨量点雨量为 321 mm (见图 3), 比实况 491 mm 小。事后反查, 已知解 8506 (与 9015 一样, 在浙江坎门登陆) 的最大雨量点的雨量为 425 mm, 与 9015 更近。这说明: 1) 降水成因太复杂。9015 是秋台风, 8506 是盛夏台风, 气候背景和天气形势均不同, 故 8506 台风不能成为 9015 台风的解元素是可以理解的。但是降水成因复杂到这种程度, 以致尽管大相径庭的几个变量值, 也可能会有相近的降水结果。2) 重解定义太宽。实际上在温州—泉州之间登陆的台风都可能造成 1 到 5 的降水, 但本文只用数码 5 代表上述登陆地段; 因此, 有可能将差的已知解选上, 将好的已知解剔除。

(2) 变量预报能力有限

9012 台风路径特殊, 在厦门附近作三次登陆, 尔后减弱消失。它比通常在厦门登陆的台风所带来的雨量大、范围广。9012 有两个雨量中心, 一个在浙南闽北, 一个在闽南 (见图 8)。本例报小了前者, 漏报了后者 (见图 7、图 8)。原因可能是: 1) 台风滞留少动, 延长了降水时间; 而目前尚未找到直接反映台风滞留少动的变量。2) 不少学者对台风路径右侧的降水 (即台风倒槽降水) 作了深入的天气学分析, 而至今对台风路径左侧的降水研究较少; 本文的变量也着重前者。还有, 28°N 以南记录少也是重要原因。

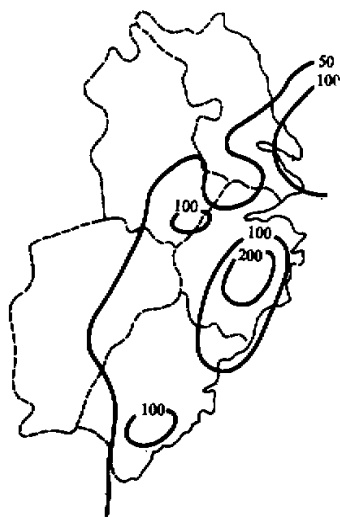


图 7 9012 台风的预报集成

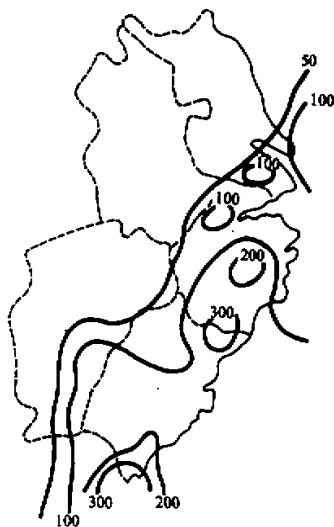


图 8 9012 台风的雨量实况

参 考 文 献

- [1] 邱崇践、丑纪范, 1988, 预报模式识别的扰动方法, 大气科学, 12, , No.3, 225—232.
- [2] 刘式达、刘式适, 1989, 非线性动力学和复杂现象, 114 页, 气象出版社.
- [3] 于禄、郝柏林, 1984, 相变和临界现象, 88 页, 科学出版社.
- [4] 贡九鼎、李慧武, 1990, 辗转坐标筛选, 大气科学, 14, No.3, 328—334.

Forecast Epitomized of Typhoon's Precipitation in East China

Gong Jiuding, Niu Xuexing and Li Huiwu

(Zhejiang Meteorological Bureau, Hangzhou 310021)

Abstract

It is an urgent need to improve forecast epitomized in the present operational forecast. The paper defines a double solution for typhoon's precipitation. According to double solution, an automatic screening method is used to process the predicted results. Through a man-machine interactive interface, we make a forecast epitomized from a few outcomes which are far from the mean but apt to heavy typhoon's precipitation forecast.

Key words: consensus; typhoon; coincidence resolution; similar.