

夏季分县降水量集成预报试验^①

苗爱梅 郭玉玺 梁明珠 袁怀亭 贾利冬

(山西省气象局, 太原 030002)

提 要

将国家气象中心 T106 和 HLAFS、山西省气象台 BP 神经元和动力相似方法输出的降水预报值作为预报因子,运用灰色理论分别将以上 4 个预报因子原始数列和降水实况数列作一次累加生成处理,分别得到随机性被弱化的单增数列,用卡尔曼滤波法进行递推计算,得到降水预报值。将得到的降水预报值作还原处理,即累减生成后,最终输出降水量分县预报。此方法用于 1998 年夏季,取得了良好的预报效果。

关键词: 数值预报产品 降水量 分县预报 灰色理论 卡尔曼滤波

引 言

目前,数值预报产品的释用已成为各级气象台站预报研究的重大课题。就降水预报来说,采用不同的方法解释应用,就会有不同的预报结果生成。如何将各种释用方法所得到的预报结果作很好的集成预报,使其预报准确率高于各种独立预报方法,当前还没有一种较完善的方法。但出于预报业务现代化的需要,笔者在 1998 年夏季降水中进行了初步尝试,并取得了较满意的效果。

1 降水集成预报方法原理

作为一个预报员,首先是要从大的形势场考虑有无降水形成,其次要考虑有哪些影响系统,在这些系统影响下,本预报责任区内哪些区域、哪些站点可能有降水,降水量级有多大。针对不同的数值产品,面对不同的预报方法所做出的多种预报结果,预报员最大的可能是相信近期预报效果好的预报方法所输出的预报结果,并依此为参考来形成自己的预报结论。因此,降水集成预报系统首先要从大的形势场判断有无降水形成。其次,要想办法使集成方程系数发生动态变化,使近期预报效果好的预报方法在集成方程中得以有较

大的权重系数。但就目前来讲,只有卡尔曼滤波能使方程中的系数发生动态变化。然而,卡尔曼滤波法又只适用于连续性数据序列。因此,要想办法将降水量这种不连续数列变为近似的连续性数据序列,尔后再用卡尔曼滤波递推计算,这就导出了对原始数据序列作一次且仅作一次累加生成处理的问题。

2 具体做法

2.1 用聚类分析将形势场分型

取国家气象中心 1995~1997 年 6~8 月的 T106 数值产品资料中的 500hPa、700hPa、地面三层形势分析场和与其相对应时期的降水实况资料进行聚类分析,将形势场分为有降水和无降水两类。

格点资料的处理范围为 30~50°N、95~125°E,网格距为 1°×1°,资料年限为三年,时次为 08 时、20 时两个时次。相应的降水实况资料为山西省 59 个站的逐日雨量。采用的聚类方法为系统聚类法(系统树)。

2.1.1 系统聚类

在聚类之前,作为样本的 500hPa、700hPa、地面分别对应所选的样本资料各自为一类,每进行一步将距离最小的两类合并

^① “九五”配套预报方法研究课题资助

成一类;并类后,计算新类与其它类的距离,构成新的距离矩阵。新类之间的距离有多种不同的定义方法,集成预报试验中采用的是类平均法。由于本系统只将形势场分为有降水和无降水两类,因此,程序设计时当类别 > 2 时,继续聚类,当类别 ≤ 2 时,聚类结束。聚类终止后求出各类的平均场,将每一层(500hPa、700hPa、地面)的 2 个平均高度场(或气压场)作为有雨、无雨的标准模型场。

2.1.2 样本矩阵

设某一数值预报产品格点报在所取地域范围每份资料(每个样本)有 m 个格点,对 n 份资料(n 个样本)来说,可以得到一组样本矩阵 X ,记为

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1j} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2j} & \cdots & X_{2m} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ X_{i1} & X_{i2} & \cdots & X_{ij} & \cdots & X_{im} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nj} & \cdots & X_{nm} \end{bmatrix}$$

在矩阵 X 中,每一行为一个样本,即一份有 m 个格点的数值产品资料,例如第 i 个样本即为 $X_{i1}, X_{i2}, \dots, X_{im}$ 。而矩阵中每一列为一个指标,即对应为某一固定序号的格点值。对于 n 份资料来说,即为第 j 个格点值。显然,每份资料(每个样本)的特征都可用相应的 m 个格点值(指标)来描述,而每个样本又看作 n 维空间的一个点。

2.1.3 距离系数(简称距离)

样本可视为 n 维空间中的点,样本间的距离也就是以各种形式定义的点与点之间的距离。

设 d_{ij} 为第 i 个样本与第 j 个样本之间的距离,且有

- (1) $d_{ij} = 0$ 时,样本 i 与 j 恒等;
- (2) 对任何 i 与 $j, d_{ij} \geq 0$;
- (3) $d_{ij} = d_{ji}$;
- (4) $d_{ij} < d_{ik} + d_{kj}$ 。

由常用的欧氏距离公式

$$d_{ij} = \sqrt{\sum_{k=1}^m (X_{ik} - X_{jk})^2}$$

($i, j = 1, 2, \dots, n$)

可知,若两个样本越相似,则它们之间的距离越小。在聚类分析中,总是把两个距离最小的样本归为一类。

2.2 预报因子的选取和插值处理

选用国家气象中心 T106 和 HLAFS、山西省气象台 BP 神经元和动力相似方法输出的降水量预报值作为预报因子,并将 T106 和 HLAFS $1^\circ \times 1^\circ$ 的格点资料(降水量预报值)插值到 59 个站点上。样本长度 $n = 60$,资料处理范围为: $32 \sim 43^\circ N, 108 \sim 115^\circ E$ 。插值处理采用距离权重法,其插值公式为:

$$R(N, E) = \frac{1}{3} \sum_{i=1}^4 (L - M(i)) R(i)$$

$$\text{其中 } L = \sum_{i=1}^4 M(i);$$

$$M^2(i) = X^2(i) + Y^2(i);$$

$$X(i) = (E - E(i)) \times R \times \pi \cos N(i) / 180;$$

$$Y(i) = (N - N(i)) \times R \times \pi / 180$$

式中, $R(N, E)$ 为某一测站的插值结果, N, E 分别为该测站的纬度和经度, $R(i) i = 1, \dots, 4$ 为距离测站最近的 4 个格点的纬度和经度。T106 和 HLAFS 模式降水预报用该插值方案进行插值。

2.3 用灰色理论对原始数列作累加生成处理

灰色理论视不确定量为灰色量。而灰色预测则不是把观测数据序列视为一个随机过程,而是看作随时间变化的灰色量或灰色过程,通过累加生成和累减生成逐步使灰色量白化,从而建立相应于微分方程解的模型并作出预报。本试验中由于预报因子用的是各种模式或方法预报的降水量值,因此,不做关联度分析。降水量预报也只是在数据的预处理和后处理上应用了灰色理论,建模和递推计算则采用了卡尔曼滤波法。用灰色理论对原始数据序列作预处理。

令 $\{X_k^{(0)}(i)\}, \{Y^{(0)}(i)\} (k = 1, 2, \dots, m; i = 1, 2, \dots, n)$ 分别为预报因子和预报量的原始数列,对其作一次累加生成:

$$X_k^{(1)}(i) = \sum_{j=1}^i X_k^{(0)}(j)$$

$$Y^{(1)}(i) = \sum_{j=1}^i Y^{(0)}(j)$$

得到新的数据序列 $\{X_k^{(1)}(i)\}$ 、 $\{Y^{(1)}(i)\}$ ，生成后的序列随机性被弱化，变成单增数列。这里 $m = 4$ 为预报因子数， $n = 60$ 为序列长度。

$$W \approx \begin{bmatrix} (\Delta\beta_1)^2/\Delta T & 0 & \cdots & 0 \\ 0 & (\Delta\beta_2)^2/\Delta T & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & (\Delta\beta_m)^2/\Delta T \end{bmatrix}$$

其中 $\Delta T = 30$ 。

$$V = \begin{bmatrix} g_1/(k - m - 1) & 0 & \cdots & 0 \\ 0 & g_2/(k - m - 1) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & g_n/(k - m - 1) \end{bmatrix}$$

式中 g_1, g_2, \dots, g_n 为预报量 $Y^{(1)}(i)$ 的 n 个残差， k 为样本容量（这里取 60）， m 是因子个数（这里 $m = 4$ ）。

2.4.2 递推计算

在确定了起步参数、建立了所需的数据文件后，将这些数据读入递推滤波系统，递推系统就可根据下面给出的 6 个递推滤波公式：

$$\hat{Y}_t = X'_t \beta_{t-1} \quad (1)$$

$$R'_t = C'_{t-1} + W \quad (2)$$

$$\sigma'_t = X'_t R'_t X'^T_t + V \quad (3)$$

$$A'_t = R'_t X'^T_t (\sigma'_t)^{-1} \quad (4)$$

$$\beta_t = \beta_{t-1} + A'_t (Y'_t - \hat{Y}_t) \quad (5)$$

$$C'_t = R'_t - A'_t \sigma'_t A'^T_t \quad (6)$$

进行递推计算。

这里需要指出的是： \hat{Y}_t 是递推系统在 $t - \Delta t$ 时刻输出的 t 时刻的预报值，此值没有经过累减，直接存入 t 时刻预报值 \hat{Y}_t 文件中，并与 t 时刻站点实测资料 Y_t 经过一次累加生成处理后的值 Y'_t ，相减生成预报误差数据文件供下一时刻递推计算用。特别强调：以上 6 个方程中，除 W 、 V 在确定后不发生变化外，其它均在发生变化，且都是经一次累加生成

2.4 用卡尔曼滤波法进行递推计算

2.4.1 递推系统参数初值的计算

对作过一次累加生成处理后的预报因子 $X_k^{(1)}(i)$ 和预报量 $Y^{(1)}(i)$ ，按照通常求回归系数估计值方法求得 $\hat{\beta}_0$ ，假定 β_0 与理论值相等， C_0 即被确定为 m 阶零方阵。

处理后的值。如： \hat{Y}_t 及 X'_t 分别为经过一次累加生成处理后的预报值和预报因子， β_{t-1} 为累加生成处理后的预报量和预报因子所产生的方程系数估算值， R'_t 为 t 时刻 β 外推值的误差方差阵， C'_{t-1} 为滤波值 β_{t-1} 的误差方差阵， W 是动态噪声的方差阵， σ'_t 是预报误差方差阵， X'^T_t 为预报因子 X'_t 的转置矩阵， V 是量测噪声的方差阵， A'_t 是增益矩阵， $(\sigma'_t)^{-1}$ 是 σ'_t 的逆矩阵。式 (5) 为系数 β_t 的订正方程， Y'_t 是 Y_t 经过一次累加生成后的值， C'_t 为 t 时刻的误差方差。

3 预报流程

图 1 为降水集成预报系统的预报流程图。实时运行时，直接从 NT 网上调用各种数值预报产品的格点资料进行解读。将解码和范围选取处理后的 500hPa、700hPa 和地面三层格点资料读入聚类分析器进行聚类分析。若三层中有两层分析为有降水则显示有降水，反之显示无降水。形势场显示有降水时，将插值处理后的 T106、HLAFS 降水预报值和 BP 神经元、动力相似法输出的降水预报值以及降水量实测值利用有限记忆法分别追加到各自的数据文件中。然后分别将各

数据文件中的数据序列作一次累加生成处理后重建另一组单增数列数据文件。将这些单增数列数据文件读入卡尔曼递推系统进行递推计算。由递推系统生成的预报结果一是追加到预报量的预报值文件中,二是对其作还原处理即累减生成处理:

$\hat{Y}(i) = \hat{Y}^{(1)}(i) - \hat{Y}^{(1)}(i-1)$, 输出最终的分县降水量预报结果。

若形势场显示无降水时不再进入递推系统计算,直接输出无降水的预报结果。

雨:0.0~9.9mm;中雨:10.0~24.9mm;大雨:25.0~49.9mm;暴雨: ≥ 50.0 mm 的评定。其预报准确率由下式计算:

$$TS = \frac{H}{F + O - H} \quad (7)$$

式中 TS 为某一量级的预报准确率; O 表示观测到的该降水量级的站次数; F 表示预报该降水量级的站次数; H 表示该降水量级预报和实况一致的站次数。

4.1 一般性降水预报效果

按照式(1)分量级统计一般性降水(小雨、中雨、大雨)的预报准确率 TS 值,结果见表1。

表1 一般性降水分县预报准确率

	观测站次数 O	预报站次数 F	正确站次数 H	预报准确率 TS
小雨	1801	1780	1578	0.788
中雨	803	701	500	0.498
大雨	78	52	43	0.494

表1显示,该降水集成预报系统对一般性降水有较好的预报效果,其中小雨的预报准确率最高,达0.788;比动力相似法、HLAFS、BP神经元、T106的预报准确率分别高出0.39、0.31、0.29、0.28;中雨的预报准确率为0.498;分别比HLAFS、动力相似法、BP神经元、T106的预报准确率高出0.15、0.13、0.11、0.10;大雨的预报相对稍差,但仍比BP神经、HLAFS、动力相似法、T106的预报准确率高出0.20、0.18、0.17、0.15。

4.2 暴雨预报效果

对暴雨预报准确率的统计从两方面来进行,即暴雨落区预报和暴雨落点预报统计。暴雨落区预报是将山西省分为7片。大同、朔州为一片;晋中为一片(含太原和阳泉);长治、晋城为一片;运城、临汾、吕梁、忻州各自为一片。暴雨落区预报的预报量和实况值分别取每片中测站降水预报和雨量实况的最大值,预报准确率仍按式(7)计算,结果列于表2。

由表2可知,1998年夏季出现的暴雨,16片次暴雨有15片次预报正确;35站次单站暴雨中有20站次预报正确。暴雨落区和落点预报准确率分别达到0.714和0.425。这表明:降水集成预报系统对夏季暴雨落区和落点都有较强的预报能力,其预报质量明显高于主观经验预报和各种独立预报方法和模

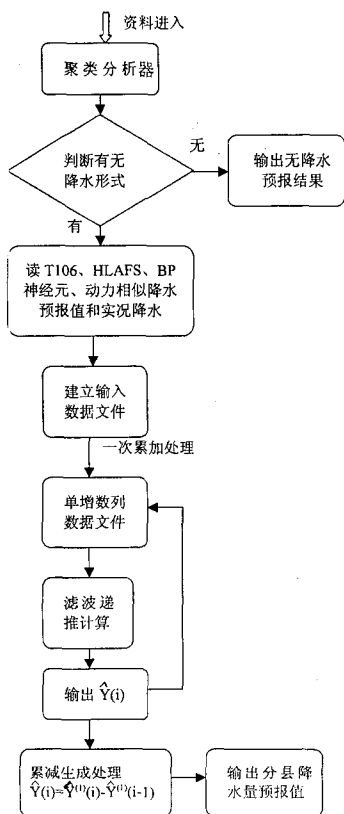


图1 降水集成预报流程

4 预报试验效果

利用该降水集成预报系统对1998年6月3日~8月31日90天进行连续运行,预报时效分为12小时、24小时、36小时、48小时4个时段。降水量评定时,我们只作了24小时一个时段、降水量划分为4个等级:小

式水平,尤其是对 50mm 以上的强降水预报显示出良好的业务使用价值。

表2 暴雨预报准确率

	观测次数	预报次数	正确次数	预报准确率
	<i>O</i>	<i>F</i>	<i>H</i>	<i>TS</i>
区域	16	20	15	0.714
落点	35	32	20	0.425

4.3 一次成功的大暴雨预报

实时运行中,有很多成功的个例,其中以1998年7月8日山西南部的大暴雨预报最为典型。这次强降水是1998年夏季降水强度最强的一次过程。表3给出了这次过程降水集成预报系统预报的部分结果和对应的降水实况。

表3 1998年7月8日大暴雨降水集成预报系统预报情况 (单位:mm)

站名	7日预报	8日实况	站名	7日预报	8日实况
安泽	79.4	67.0	襄垣	100.1	89.0
吉县	48.0	52.0	万荣	71.0	58.0
长治	71.0	54.0	运城	55.2	30.0
平顺	25.0	33.0	垣曲	29.1	19.0
榆社	18.4	27.0	沁源	50.2	64.0
侯马	101.2	135.0	芮城	10.1	33.0
翼城	53.1	68.0	平陆	9.9	26.0
阳城	27.1	31.0	沁县	38.1	46.0
晋城	10.0	21.0	霍县	42.5	52.0

由表3可知:这次强降水过程落区和落点预报都很成功,在本省的西南和东南各预

报有一个强降水中心,在所列的18个站点上,预报50mm以上降水的有8个站点。实况:两个强降水中心预报准确,18个站点中共有9个站降水在50mm以上,1个站在100mm以上。本次降水暴雨的落点预报评定:空报1站、漏报2站,预报正确站次为7站,暴雨落点*TS*评分高达0.70。

5 结语

①本试验结果表明,该降水集成预报系统模拟了值班预报员的全程预报思路。尤其是将降水量这种不连续数列经过一次累加生成处理变为单增数列,而后用卡尔曼滤波递推计算,使集成方程中的系数发生动态变化,达到近期预报效果好的方法或模式得以较大的权重系数,最终使预报量的绝对误差减小,预报质量提高。

②该集成系统的特点一是不需要长的时间序列,不受资料年代的限制,解决了目前数值预报模式不断更新换代的问题;二是它在资料处理上采用了有限记忆法,其样本逐日更新为一定序列长度的前期资料,既消除了季节影响,又使递推计算灵敏度提高。

③用聚类分析将形势场分为有降水和无降水两类,有效地控制了空报现象和过程结束时预报不够理想的弊端。

Composite Prediction Experiment of County Rainfall in Summer with NWP Products

Miao Aimei Guo Yuxi Liang Mingzhu Yuan Huaiting Jia Lidong

(Shanxi Meteorological Bureau, Taiyuan 030002)

Abstract

The rainfall predicting values from T106 and HLAFS of NWC, BP neuron and dynamic similar method of Shanxi Meteorological Observatory were used for prediction factors. By using of grey theory, the original sequence of the predicting factors and actual rainfall sequence were conducted for accumulating process, respectively, then the random weakened increasing sequence were obtained. The recurrence calculation with Kalman filter method was used to acquire the rainfall predicting values. The values were retrieved to initial state, and the rainfall prediction for every county was given. The experiment in Summer in 1998 was conducted and obtained a good result in rainfall prediction.

Key Words: NWP product rainfall prediction for every county grey theory Kalman filter