多层递推平稳模型

汤兆焘　李法然　杨育强　黄玲琳　孙建明

（浙江省湖州市气象局，313000）

提　要

通过对小概率事件历史样本作多层递推迭代，逐次改变引入样本数求取权函数，建立一种平稳统计模式。模式信息容量大，收敛快，克服了常规统计模式的局限性，在灾害性天气客观预报模型应用中取得了明显效果。

关键词： 小概率事件　环流背景　多层递推

前　言

常规的数理统计模型已在气象领域中得到广泛的应用，然而针对灾害性天气等小概率事件，尽管对统计模型进行了改进，但预报模型往往出现拟合率高，业务预报准确率低的问题。究其原因，我们认为主要由于各类统计模型所获取的信息量小，而不能全面描述小概率事件的环流背景。近几年来，人们试图利用人工智能，建立各种类似专家系统的推理模式，其必备的规则（或知识）大多数取自专家的经验，由于规则的不完备、不客观性以及推理过程计算误差等等，往往导致推理失败，更何况推理系统中的学习机制难于有效地建立，因此离实际业务应用尚有一定的差距。我们提出多层递推平稳模型，试图解决灾害性天气的预报问题。

1　模型的技术思路

设小概率事件为 B，必然存在引发 B 的天气背景场 A，也就是说导致 B 的发生，必须具备某种特定的条件。其中，A 包括天气环流背景、水汽条件、层结稳定度及各种触发机制。显而易见，A 是 B 的必要条件，而并非充要条件。这是因为人们对小概率事件的认识还不够全面，且所拥有的信息具有一定的局限性，因此条件概率：$P(B/A)$ 必然小于 1。我们试图寻找最为完备的 A 场，用来描述事件 B。此时，当出现 A 则发生 B 的条件概率 $P(B/A)$ 达到最大，这正是多层递推平稳模型的设计目的。我们引入足量的天气信息，从中计算出 A 场，从而建立相应的预报模式。

1.1　模型概况

现有预报量 Y，其中尽可能多地包含 B 类天气的历史事件，同时引进占有一定比例的不同季节的 B 的逆事件 $\bar{B}$ 个例，如引入暴雨样本以建立暴雨预报模型时，按某种比例同时引进大雨、中雨、小雨和无雨的样本，使 Y 呈正态分布，以便提高预报模型的灵敏度。我们首先根据天气的剧烈程度来确定预报量的初值，通过标准化处理后，使 Y 处于 0，1 之间，此时的 Y 仅作为模型的猜值，即 Y 的初始量。

设已拥有各类线性或非线性的天气因子，通过相关普查初选因子，再经独立性检验和标准化处理，并通过变换使因子和 Y 呈正

---

* 本文执笔者.

相关，从而形成因子集 X，$X\in[0,1]$。

$$X=\begin{bmatrix}x_{11} & x_{12} & \cdots & x_{1p}\\ x_{21} & x_{22} & \cdots & x_{2p}\\ \vdots & \vdots & & \vdots\\ x_{n1} & x_{n2} & \cdots & x_{np}\end{bmatrix}=(X_1,X_2,\cdots X_p)$$

其中 n 为总体样本，p 为模型因子总数，$X_i=(x_{1i},x_{2i},\cdots x_{ni})^T$。我们将计算出一组因子权重系数 W，形成了多层递推模型：

$$Y=\frac{1}{p}WX+\varepsilon$$

其中，W 为因子的权重系数，表征因子与 Y 的关联程度，即描述因子对 Y 的贡献，它取决于因子的类型、地理位置、季节性差异等。我们以因子与 Y 之间的相关系数来代替两者之间的权重系数，即：

$$W=r=S_Y^{-1/2}*S*S_X^{-1/2}$$

ε 为随机误差，由于 Y 呈正态分布，因此：

$$\varepsilon=(\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i-y_i)^2)^{1/2}$$

多层递推模型的准则是使 ε 趋于极小值，即满足：

$$\lim_{k\to M}\frac{\delta\varepsilon}{\delta k}=0$$

这里，M 为足够大的迭代次数。

1.2 模型建立过程

第一层迭代，即取 $k=1$。设总样本数为 N，首先进行 Y 量赋值：$Y^{(k)}=Y$。

①取起始样本 L 和滑动步长$\triangle L$。

一般取 L 和$\triangle L$ 在 $N/10$—$N/8$ 之间；计算出递推次数：

$$T=\mathrm{INT}[(N-L)/\triangle L]+1$$

②计算 L 个样本的因子的权重系数及 Y 的估计值 $\hat{Y}^{(k)}$：

$$W_{(L)}=S_{Y_{(L)}^{(k)}}^{-1/2}*S_{(L)}*S_{X_{(L)}}^{-1/2}$$

$$\hat{Y}_{(L)}^{(k)}=\frac{1}{p}\sum_{j=1}^{p}W_{(L)j}*X_j$$

这里，$\hat{Y}_{(L)}^{(k)}$ 为 $Y_{(L)}^{(k)}$ 在因子场中的映射。

③增加$\triangle L$ 样本量，并以 $\hat{Y}_{(L)}^{(k)}$ 替代 $Y_{(L)}^{(k)}$ 参加计算。此时：

$$Y_{(L+\triangle L)}^{(k)}=[\hat{y}_{(1)}^{(k)}\cdots\hat{y}_{(L)}^{(k)},y_{(L+1)}^{(k)},\cdots y_{(L+\triangle L)}^{(k)}]$$

④计算 $W_{(L+\triangle L)}$ 及 $\hat{Y}_{(L+\triangle L)}^{(k)}$

$$W_{(L+\triangle L)}=S_{Y_{(L+\triangle L)}^{(k)}}^{-1/2}*S_{(L+\triangle L)}*S_{X_{(L+\triangle L)}}^{-1/2}$$

$$\hat{Y}_{(L+\triangle L)}^{(k)}=\frac{1}{p}\sum_{j=1}^{p}W_{(L+\triangle L)j}*X_j$$

逐次增加 ΔL 样本量，重复③④步计算，直到引入全部样本，并计算结果。

⑤最后计算均方差：$\varepsilon^{(k)}$

$$\varepsilon^{(k)}=[\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i^{(k)}-y_i)^2]^{1/2}$$

至此完成了第一层递推计算。如需继续 k 层递推，则首先进行 Y 量赋值，即：

$$Y^{(k)}=\hat{Y}^{(k-1)}$$

并重复①—⑤步，若 $\varepsilon^{(k)}-\varepsilon^{(k-1)}\leqslant\lambda$，则完成整个递推过程，其中 λ 为精度值，此时，$Y^{(k)}=\frac{1}{p}WX$ 为业务预报模型。$Y^{(k)}\in[0,1]$。建模流程图略。

1.3 举例说明

1.3.1 样本选取及处理

我们选取 1980—1992 年的 3—10 月暴雨个例，以及一定比例的晴雨天气样本。总样本 N 为 111 个，其中暴雨样本 55 个，一般性降雨样本 33 个，无雨样本 23 个，最大日雨量为 188mm。经标准化处理，使预报量 $Y\in[0,1]$。如 $Y=1$，即代表 188mm 雨量。

1.3.2 因子选取

选自高空三层，即 850、700、500hPa 站点资料，计算各类物理量，经气候平滑、标准化处理，通过相关普查、自相关剔除，形成预报因子集。其中包含水汽输送、K 指数、垂直速度等 67 个因子进入暴雨天气预报模型。

1.3.3 建立预报模型

取 $L=\Delta L=10$，计算得：$T=11$，选取 $\lambda=2*10^{-4}$作为递推阈值。

第一层迭代，即取 $k=1$，首先进行 Y 量赋值：$Y^{(1)}=Y$。

引入 10 个样本，计算出 $W_{(10)}$ 和 $\hat{Y}_{(10)}^{(1)}$

增加 ΔL 样本，且以 $\hat{Y}_{(10)}^{(1)}$ 替代 $Y_{(10)}^{(1)}$，此

时：

$$Y_{(20)}^{(1)} = [\hat{y}_{(1)}^{(1)}\cdots\hat{y}_{(10)}^{(1)}, y_{(11)}^{(1)}\cdots y_{(20)}^{(1)}]$$

计算得：$W_{(20)}$ 及 $\hat{Y}_{(20)}^{(1)}$，逐次增加 ΔL 样本，直到引入全部样本，计算得：$W_{(111)}$ 及 $\hat{Y}_{(111)}^{(1)}$

最后计算 $\epsilon^{(1)}=[\frac{1}{111}\sum_{i=1}^{111}(\hat{y}_i^{(1)}-y_i)^2]^{1/2}=0.1107245$。

由于 $\epsilon^{(1)}>\lambda$，即进入下一层递推。

第二层递推：即 $k=2$，首先进行 Y 量赋值：$Y^{(2)}=\hat{Y}^{(1)}$。

引入 10 个样本，计算出 $W_{(10)}$ 和 $\hat{Y}_{(10)}^{(2)}$，以 $\hat{Y}_{(10)}^{(2)}$ 替代 $Y_{(10)}^{(2)}$ 参加计算。

此时：$Y_{(20)}^{(2)}=[\hat{y}_{(1)}^{(2)}\cdots\hat{y}_{(10)}^{(2)}, y_{(11)}^{(2)}\cdots y_{(20)}^{(2)}]$

计算得：$W_{(20)}$ 及 $\hat{Y}_{(20)}^{(2)}$，再替代 Y 量，增加样本，重复上述计算，最后计算得到：$W_{(111)}$ 及 $\hat{Y}_{(111)}^{(2)}$，并计算

$$\epsilon^{(2)} = [\frac{1}{111}\sum_{i=1}^{111}(\hat{y}_i^{(2)} - y_i)^2]^{1/2} = 0.1105938$$

此时，$\epsilon^{(2)}-\epsilon^{(1)}=0.0001307\leqslant\lambda$，结束递推。

由此形成暴雨预报模型：$\hat{Y}=\frac{1}{p}WX$，暴雨临界值 $Yc=0.28$，如 $Y>Yc$，则起报暴雨。

拟合情况：报对暴雨 36 次，空报 15 次，漏 19 次。成功界限指数 0.56，拟合率 77/111＝79.4％。

2 模型特点

2.1 该模型利用预报量猜值 Y 进行多层递推迭代，每次迭代结果 Y 替代上一步计算的 Y 量，从而避免预报量初值确定的人为性，客观地描述了天气背景场特征。在常规统计模型中，模型的准确性在很大程度上依赖于预报量 Y，而 Y 完全由人为确定，特别对灾害性天气其突发性、局部性特点，很难进行客观描述，这是影响预报实用性的一个方面。

2.2 该模型因子信息量大，一般可取 50—100 个因子，如果计算容量允许还可更多。因子间互补能力强，能较全面地反映各种灾害性天气的环流背景和各种触发机制。针对某一类型的灾害性天气，它是环流背景、大气能量、水汽条件、各类触发机制等众多因素综合影响的结果，仅用几个因子往往很难对其进行确切描述。因此，常规统计模型由于引入信息量小，难于建立实用的预报模型。

2.3 该模型计算过程收敛性好。由于模型的理论基础是寻找能反映天气的背景场 A，并进行多层逐次迭代，因此，随着迭代次数的增加，因子权重系数很快趋于稳定，收敛迅速。

2.4 模型稳健性好，预报质量高。任一统计方法的优良性往往依赖于我们对样本的分布的假设，我们对样本进行处理，使某一特定的天气属于正态分布。但有时这些基本假设完全可能被破坏。实际上，象灾害性天气这样的小概率事件恰恰为非正态分布，一般而言，在很大程度上会直接影响模式的预报质量，造成实际预报质量低下。然而，我们用多层递推平稳模型建立的暴雨和强对流两个预报系统，1993 年 5—7 月投入实际业务使用，效果十分明显。其中，暴雨预报系统起报大一暴雨 9 次，报对 8 次，空漏报各 1 次，成功界限指数为 8/10＝0.8；强对流预报系统：起报短时暴雨 9 次，报对 7 次，准确率为 7/9＝77.9％，起报大风 5 次，报对 4 次，准确率为 4/5＝80％。

3 建立预报模型有关技术处理

3.1 引入样本合理化。由于灾害性天气具有偏态特征，为了提高模型的稳定性，应适量引入其它类型，如晴雨的样本。当预报对象存在不同气候特征时，应引入同季节的一般天气样本，以提高对灾害性天气的分辨率。

3.2 消除气候差异。在因子处理过程中，由于因子本身存在气候性差异，因此需采用相应技术去除其气候脉动，例如，大气能量春季小、夏季大，相差可达几倍；并且，春季高能时常常发生剧烈天气，而夏季则不尽然。因此，我们用地面能量作为参照体，进行客观处理。经处理后的因子去除了气候性差异。

3.3 由于模型采用逐次递推迭代，原则上要求样本的排序应达到各类天气相互交叉分

布，这有利于快速收敛，否则，计算时收敛速度较慢。

3.4 整个过程采用因子系数矩阵自动传递技术，因此具有很强的业务应用能力。

4 讨论

由于目前人们对灾害性天气等小概率事件的物理机制认识不够全面深入，而此类事件的发生发展机制极其复杂，因此还难以较准确地用数理模式进行模拟，只能停留在统计方法上。本文从数理统计角度出发，客观地寻找其背景场，并进行多层逐次递推，尝试建立适合于全年的预报模式。在实际业务应用中已发挥积极的作用，其预报质量接近或超过经验预报水平。尽管各模型的使用效果较好，但对于适用于全年各个不同季节的预报模式，还存在少量的空报现象。我们归纳其原因如下：

4.1 我们所用资料局限于 500hPa 以下，而剧烈天气往往可发展到对流层顶；另外，没有进行地形条件的数值化处理，缺乏下垫面的某些信息。这些重要信息的缺少，影响了描述天气背景的客观性。

4.2 我们采用参照体及要素时间函数来消除气候差异，这些处理技术还不够完备，因此在盛夏季节副高控制的高温高湿天气条件下会出现部分空报现象，这是影响总体预报质量的一个主要因素。

参考文献(略)

A Multi-Level Recurrence Equable Model

Tang Zhaotao Li Faran Yang Yuqiang Huang Linglin Sun Jianming

(Huzhou Meteorological Bureau, Zhejiang Province, 313000)

Abstract

Through Multi-level recurrence iteration for historical samples of small-probability events, inducted sample numbers are progressively varied to win weight function. An equable statistical model is developed with this method. The model contained more information and converged faster. It overcomes the limitation of annual statistical model. This has obviously increased the capability in application of the objective forecasting model to the calamitous weather.

Key Words: small-probability events atmospheric circulation background multi-levels recurrence