试用非线性预报方法预报辽宁夏季气温和降水

冯耀煌
（沈阳区域气象中心研究所）

夏梅艳
（沈阳中心气象台）

一、前　言

过去在长期预报的统计分析中大部分采用相关分析方法，找出相关较好的关键区，如要做定量的预报，则找出相关较好的预报因子，用逐步回归的方法建立预报方程，但是这里的相关分析及建立的逐步回归预报方程都是线性的，而实际上预报量与预报因子的关系是非线性的。因此，线性相关分析方法是不适合实际的，用线性预报因子建立起来的预报方程也必然会影响预报的准确率，本文就是为了克服这个缺点，采用优化的非线性预报方法，首先普查出最优的非线性预报因子，然后用限定重要因子进入方程的逐步回归方法找出最优的非线性预报方程，并试用这方法对辽宁夏季气温和降水进行了预报，取得了比线性预报方法较好的结果。

二、方法概述

非线性预报方法在参考文献〔1、2、4〕中有较详细介绍，现把其基本内容概述如下：

（一）利用最优化方法普查因子，确定因子的类型

设预报量$Y=\{y_1, y_2, \cdots\cdots y_N\}$，对应预报因子$X=\{x_1, x_2 \cdots\cdots x_N\}$，如何衡量预报因子的好坏呢？以前一般采用计算线性相关系数办法，因Y与X的相关系数的大小反映了变换后点集$\{(x'_i\ y'_i), i=1, \cdots N\}$在$Y'=X'$直线附近的情况〔2〕，相关系数大就说明所有这些点离直线$Y'=X'$比较近，否则就比较远，现在推广到非线性情况，即点集可在$Y=F(X)$曲线附近摆动，目前假设$F(X)$有以下4种类型：

1.线性函数类型：$Y=X$。2.幂函数类型：$Y=X^a$。3.指数函数类型：$Y=e^{ax}$。4.对数函数类型：$Y=\ln X$。

对于某个具体因子究竟属于哪种类型，还要经过计算相关系数，然后比较来决定，计算相关系数公式如下：

$$R=\frac{\sum_{i=1}^{N}(x'_i-\overline{x}')(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{N}(x'_i-\overline{x}')^2\sum_{i=1}^{N}(y_i-\overline{y})^2}}$$

计算第1种类型时，$x'_i=x_i$，$\overline{x}'=\frac{1}{N}\sum_{i=1}^{N}x_i$；计算第2种类型时，$x'_i=x_i^a$，$\overline{x}'=\frac{1}{N}\sum_{i=1}^{N}x_i^a$；

计算第 3 种类型时，$x_i' = e^{ax_i}$，$\overline{x}' = \frac{1}{N}\sum_{i=1}^{N} e^{ax_i}$；计算第 4 种类型时，$x_i' = \ln x_i$，$\overline{x}' = \frac{1}{N}\sum_{i=1}^{N} \ln x_i$。这样第 1、4 种类型可直接计算得 $|R_1|$、$|R_4|$，第 2、3 种类型还带有一个参数a，我们采用最优化方法[3]，求出使相关系数最大的a 值及目标函数 $|R_2|$、$|R_3|$，然后进行比较，若 $|R_g| = \max\{|R_1|, |R_2|, |R_3|, |R_4|\}$，则预报因子属于第g 种类型。

（二）对预报因子作显著性和稳定性检验，选出最优的预报因子

在确定因子类型以后，还要通过检验，首先看是否通过相关系数显著性检验，如果通过，则再做样品和a 值的稳定性检验[1]。把通过显著性检验和稳定性检验的因子，称为最优预报因子，作为建立方程时的待选因子。把具有明显物理意义的且显著性水平较高的最优预报因子称为重要因子，这些因子在建立方程时优先考虑，将其限定进入预报方程。

（三）用多功能逐步回归办法建立预报方程，选出最优的预报方程

设在普查因子中得到m个最优预报因子X_1，X_2，……X_m，并假定前 3 个为重要因子，用一般逐步回归和至少限定一个重要因子进入方程的办法，共可建立 8 个预报方程。从中选出统计检验和试报检验都较好的作为最优预报方程，它的一般表达式为：

$$\hat{Y} = a_0 + \sum_{i\in A} a_i X_i' + \sum_{i\in B} b_i X_i$$

其中，a_0、a_i、b_i 为回归系数，X_i' 为限定进入方程的重要因子，X_i 为逐步引进的最优预报因子，A 为限定引进重要因子序号集合，B 为逐步引进的最优预报因子序号集合。

三、应用举例

我们将此方法应用于辽宁省夏季气温和降水预报，预报量Y_1为 6 — 8 月平均温度距平，Y_2为 6 — 8 月降水距平百分率，预报因子为北半球500hPa月平均高度场、太平洋海温场和球谐系数，资料从1956—1986年，试报1987年和1988年，现把计算结果叙述如下：

（一）普查因子结果

1.关键月和关键区。我们用预报量Y_1和Y_2普查了高度场、海温场及球谐系数，高度场和海温场资料为前一年 9 月至当年 2 月，高度场每个月有576个点，海温场每个月有286个点，球谐系数资料为前一年11—12月振幅和位相。现把各月份通过$\alpha = 0.01$显著性检验的因子个数统计列于表 1。

从表 1 中可看出：①有些月份通过显著性检验的因子特别多，可称为关键月 如12月的高度场和11月的海温场可说是预报量Y_1 的关键月，2 月的高度场可以说是Y_2 的关键月；②高度场和海温场对夏季气温都有明显的影响，特别是秋季和前冬（9 —12月）的高度场和海温场对夏季温度影响特别显著；③冬季（12— 2 月）高度场对夏季降水有明显的影响，而海温场对夏季降水影响较少。

表 1　各月通过显著性检验因子数

预报量＼因子场＼数、%＼月		9	10	11	12	1	2	总计
Y_1	高度场	42 7.3	44 7.6	28 4.9	70 12.2	22 3.8	14 2.4	220 6.4
	海温场	15 5.2	27 9.4	28 9.8	21 7.3	8 2.8	14 4.9	113 6.6
Y_2	高度场	10 1.7	18 3.1	11 1.9	33 5.7	23 4.0	41 7.1	136 3.9
	海温场	3 1.0	4 1.4	4 1.4	1 0.3	3 1.0	8 2.8	23 1.3

另外，在普查中还发现每个月通过显著性检验的因子还集中在一个或几个关键区附近，尤其是关键月的关键区特别明显，关键区中有时因子的类型还不一样，我们把关键区中因子的类型占多数的定为该关键区的类型。Y_1 和Y_2 的关键月中的关键区和类型如图 1 、图 2 、图 3 （略）所示。

从图 1 、图 2 看出，辽宁夏季气温与12月份的北冰洋、巴伦支海、大西洋高度场和11月份的太平洋中部、西部海温场有较明显的非线性相关。从图 3 看出，辽宁夏季降雨与当年 2 月份的印巴地区、黑海、阿拉斯加湾、美国大陆等高度场非线性相关较好。

2. 最优预报因子和重要因子。从表 1 中看出，对Y_1 来说，通过显著性检验的因子海温场有113个，高度场有220个，加上球谐系数 3 个，共336个。对于Y_2 来说海温场23个，高度场136个，加上球谐系数 7 个，共166个，从这些因子中如何选出最优预报因子呢？因大部分通过显著性检验的因子都集中在关键区附近，为了避免因子之间发生关系，我们在每个关键区选出一个相关系数高的因子作为代表。同时我们还考虑了因子的稳定性和每个月的代表性，这样Y_1 共选出了34个最优预报因子（其中高度场17个，海温场15个，球谐系数 2 个），Y_2 共选出了13个最优预报因子（其中高度场 8 个，海温场 3 个，球谐系数 2 个），然后在这些最优预报因子中，Y_1 和Y_2 分别选出 3 个稳定性程度好且相关系数高的作为重要因子，放在各个因子的前面，作为限定进入方程之用。Y_1 和Y_2 选出的预报因子如表 2 、表 3 （略）。

（二）建立预报方程结果

我们采用引进重要因子的多功能逐步回归程序，对Y_1 和Y_2 分别建立了 8 个非线性预报方程，然后从中找出一个试报误差小的作为最优回归方程，现把计算得到的一般回归方程及最优回归方程列出：

Y_1 的一般回归方程

$$\hat{Y}_1 = 19.5465 - 25.2604e^{-8.273\left(\frac{X_1-48}{19}\right)} - 29.6957e^{-0.7847\left(\frac{X_7-85}{3}\right)}$$

$$+ 25.654\left(\frac{X_{13}-85}{5}\right)^{0.137} - 33.4942\left(\frac{X_{27}-4.8}{2}\right)^{0.6497} +$$

Y_1 的最优回归方程（限定X_1、X_3进入方程）：

$$\hat{Y}_1=7.086-27.8622e^{-8.273(\frac{X_1-48}{19})}+14.5214e^{-4.0408(\frac{X_3-23.2}{3.2})}$$

$$-30.0975e^{-0.7847(\frac{X_7-85}{3})}+39.2301(\frac{X_{13}-85}{5})^{0.137}-$$

$$-36.3538(\frac{X_{27}-4.8}{2})^{0.6497}$$

其中X_1：前一年10月（50°N、130°W）500 hPa 高 度； X_3：前一年 9 月(10°S、115°W）海温；X_7：前一年 9 月（15°N、95°W）500hPa高度；X_{13}：前一年11月(15°N、135°E）500hPa高度；X_{27}：前一年12月（50°N、165°W）海温。

Y_2 的一般回归方程

$$\hat{Y}_2=-18.5622+23.3294(\frac{X_1-72}{13})^{7.4415}+23.0981(\frac{X_2-48}{26})^{3.4549}-$$

$$-12.5631\times10^{-6}(\frac{X_8-85}{3})^{-1.5052}+23.169(\frac{X_9-17.5}{2.2})^{1.383}+$$

$$+1.507e^{2.3576(\frac{X_{10}-21.9}{3.4})}$$

X_2 的最优回归方程（限定X_2、X_3进入方程）

$$\hat{Y}_2=-16.2224+20.7964(\frac{X_2-48}{26})^{3.4549}+30.5872(\frac{X_3-73}{11})^{1.3997}+$$

$$+21.3625(\frac{X_5-52}{13})^{1.3325}+23.4357(\frac{X_{11}-22.2}{3.8})^{4.0512}-$$

$$-18.8242(\frac{X_{13}-4}{8.6})^{1.2914}$$

其中,X_1：前一年10月（35°N、15°W）500hPa高度；X_2：当年 2 月（40°N、140°W）500hPa高度；X_3：当年 2 月（25°N、75°E）500hPa高度；X_5：前一年12月（40°N、30°E）500hPa高度；X_8：前一年 9 月（15°N、115°E）500hPa高度；X_9：当年 2 月（30°N、165°E）海温；X_{10}：前一年12月（20°N、115°W）海温；X_{11}：前一年9月（25°N、115°W）海温；X_{13}：前一年11月球谐系数Q_{10}^{4}。

为了对预报方程进行检验，我们假设L 为方程因子数，Q 为残差平方和，R 为复相关系数，F 为F 检验值，E R M为拟合平均绝对误差，R V 为方差缩减,FERM为试报平均绝对误差，现把Y_1、Y_2 的一般回归方程及最优回归方程检验结果列于表 4 。

从表 4 可看出：

1.无论Y_1 或Y_2，最优回归方程检验效果皆比一般回归方程要好。

2.从Y_2 检验结果看出，一般逐步回归引进方程的因子集合（X_1、X_2、X_8、X_9、X_{10}）不

表4　非线性预报方程检验结果

预报量＼项目	方　程	L	Q	R	F	ERM	RV	FERM
Y_1	一般回归	4	13.7653	0.8974	26.8844	0.5037	0.8053	0.53
	最优回归	5	13.2993	0.9010	21.5801	0.4903	0.8119	0.44
Y_2	一般回归	5	2729.8290	0.8855	18.1662	8.1003	0.7812	32.10
	最优回归	5	2588.4080	0.8918	19.4319	7.8487	0.7953	4.20

是唯一的最优子集，还可找到另一最优子集（X_2、X_3、X_5、X_{11}、X_{13}）　相应的残差平方和小于一般逐步回归的残差平方和，并且这样选出的最优回归方程不但拟合效果好，而且试报效果也好。

（三）非线性预报方程与线性预报方程比较

为了进行比较，我们用同样的预报因子用同样的方法对Y_1和Y_2分别建立了8个线性预报方程，Y_1和Y_2的最优回归方程与一般回归方程都是一样的，它们的预报方程如下：

Y_1的一般（最优）回归预报方程

$$\hat{Y}_1 = -1.0654+22.4305\left(\frac{X_7-85}{3}\right)+20.8473\left(\frac{X_{13}-85}{5}\right)-$$

$$-30.9214\left(\frac{X_{27}-4.8}{2}\right)-14.6572\left(\frac{X_{34}-2}{174}\right)$$

其中X_7：前一年9月（15°N、95°W）500hPa高度；X_{13}：前一年11月（15°N、135°E）500hPa高度；X_{27}：前一年12月（50°N、165°W）海温；X_{34}：前一年11月球谐系数Q_6^2。

Y_2的一般（最优）回归预报方程

$$\hat{Y}_2 = -19.3691+30.8598\left(\frac{X_3-73}{11}\right)+26.2063\left(\frac{X_7-82}{8}\right)+$$

$$+19.2978\left(\frac{X_{11}-22.2}{3.8}\right)-17.7058\left(\frac{X_{12}-1}{88}\right)-13.9021\left(\frac{X_{13}-4}{86}\right)$$

其中X_3：当年2月（25°N、75°E）500hPa高度；X_7：前一年11月（20°N、20°W）500 hPa高度；X_{11}：前一年9月（25°N、115°W）海温；X_{12}：前一年11月球谐系数Q_8^4；X_{13}：前一年11月球谐系数Q_{10}^4。

上面两个线性预报方程的检验结果如表5。

表5　线性预报方程检验结果

预报量＼项目	方　程	L	Q	Q	F	ERM	RV	FERM
Y_1	一般(最优)回归	4	18.7479	0.8572	18.0118	0.6377	0.7348	0.8400
Y_2	一般(最优)回归	5	3153.8650	0.8664	15.0515	8.5857	0.7506	8.6000

比较表4和表5，我们可以看到，非线性最优预报方程比线性最优预报方程效果要好得多，不但拟合效果好，而且提高了预报的精确度。

四、结束语

1.从非线性预报方法普查的结果看出，属于线性类型因子几乎是没有的，一般都是属于非线性类型，尤其是第2、3种类型最多，这样普查出来的因子相关程度高，并且符合实际，为建立最优回归方程奠定了良好基础，但是否还有其他非线性因子类型，有待今后继续研究。

2.从非线性预报方法建立方程的结果看出，用限定进入重要因子最后选最优回归方程的办法比一般逐步回归方法效果好，这样可选到拟合好、试报效果也好的预报方程。

3.从比较的结果看出，非线性预报方法比线性预报方法更加优越，不论拟合和试报上,效果都有显著的提高。

4.这种非线性预报方法，不但可以做定量预报，同样可以推广到定性预报中去（另有一套普查因子和建立判别方程的程序）。

参考文献

【1】冯耀煌，优化的非线性预报方法及其应用，现代若干动力统计分析预报方法讲习研讨会材料，1987年。

【2】冯耀煌、杨旭，最优化方法在天气预报中的应用，《气象》，（8）1987年。

【3】范鸣玉、张莹等，最优化技术基础，清华大学出版社，1982年。

【4】冯耀煌、杨旭，论最优预报因子与最优预报方程，气象学报，第47卷，(1)1989年。