第11卷 第3期 海洋与湖沼 Vol. 11, No. 3
1980年7月 OCEANOLOGIA ET LIMNOLOGIA SINICA July, 1980

考虑影响因子的经验正交预报方法的研究

苏育嵩 苏志清
(山东海洋学院海洋系)

统计预报法目前在国内外已经得到广泛利用,其特点是根据历史资料的统计规律,通过外推实现预报,这是所有统计预报法的基础。最初级的统计预报是对单站的历史资料本身的统计规律进行外推预报的[1],这种预报法既不考虑预报量之间的内在联系,又不考虑外界因子的影响,因此预报效果一般较差。改进的途径是朝着两个方向发展:或者考虑外界因子影响的单站预报;或者考虑预报量之间的内部联系的场预报[1]。前一种预报法目前使用最广,它不仅可以考虑多因子的影响,还可以对组配的因子进行挑选,从而建立各站的最佳预报方程。至于面的预报,是由各站的预报值组成的,由于各站之间并没有什么内在联系,往往导致预报效果的不甚稳定。关于后一种的场预报法,由于体现了各站之间的内在联系,目前日益引起预报单位的重视,但这种预报法一般着重于场的分解[2,3],或者与平稳时间序列法结合,它们都忽略了外界因子的影响,多数情况下,又导致预报值的过于保守。

本文的目的,是企图建立既考虑外界因子的影响,又考虑预报量间的内在联系的多因子场预报模式,同时又能对多种组配的因子进行挑选,建立最佳预报方程。

一、预报方法的基本原理

现有的利用经验正交分解法与平稳时间序列法相结合的预报量场预报方法,其最大缺点是不考虑外界因子的影响。为了把外界影响因子引入经验正交分解法的预报方程中,本文提出五种预报方法,也可将这五种预报方法归纳成三类,即确定因子的单站预报,确定因子的场预报和优选因子的场预报。现分述如下:

(一)确定因子的单站预报

在利用经验正交分解法作为预报工具中,确定因子单站预报是指因子场对单站预报量的预报。考虑同一因子的空间场(如气压场等)对各站预报量的影响的预报设想,在预报方法中是经常可以见到的,而在利用经验正交分解法时,从方法本身而论,因子场可以是同一物理量因子的空间场,也可以是不同物理量因子的组合场,这是经验正交分解法的优点之一。

1. 自回归因子场的单站预报法

首先将因子场(距平矩阵)

本刊编辑部收到稿件日期:1979年1月5日。

$$
{}_{n}X_{m}=\begin{bmatrix}\Delta X_{11} & \Delta X_{12} & \cdots & \Delta X_{1m}\\ \Delta X_{21} & \Delta X_{22} & \cdots & \Delta X_{2m}\\ \vdots & & & \\ \Delta X_{n1} & \Delta X_{n2} & \cdots & \Delta X_{nm}\end{bmatrix}
$$

进行正交分解，求得特征值矩阵 ${}_{m}\Lambda_{m}$ 与特征向量矩阵(经验正交函数) ${}_{m}\Phi_{m}$:

$$
{}_{m}\Lambda_{m}=\begin{bmatrix}\lambda_{1} & & & & \\ & \lambda_{2} & & 0 & \\ & & \lambda_{3} & & \\ & 0 & & \ddots & \\ & & & & \lambda_{m}\end{bmatrix},
$$

$$
{}_{m}\Phi_{m}=\begin{bmatrix}\varphi_{11} & \varphi_{12} & \cdots & \varphi_{1m}\\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2m}\\ \vdots & & & \\ \varphi_{m1} & \varphi_{m2} & \cdots & \varphi_{mm}\end{bmatrix},
$$

这里 n 为序列的个数，m 为因子个数 $(n>m)$，其次由 ${}_{n}X_{m}$ 及 ${}_{m}\Phi_{m}$ 求出权重系数矩阵:

$$
{}_{n}E_{m}={}_{n}X_{m}\Phi_{m}=\begin{bmatrix}e_{11} & e_{12} & \cdots & e_{1m}\\ e_{21} & e_{22} & \cdots & e_{2m}\\ \vdots & & & \\ e_{n1} & e_{n2} & \cdots & e_{nm}\end{bmatrix}, \tag{1}
$$

然后把 ${}_{n}E_{m}$ 当作新的因子(组合因子)，与预报量矩阵

$$
{}_{n}Y_{v}=\begin{bmatrix}\Delta y_{11} & \Delta y_{12} & \cdots & \Delta y_{1v}\\ \Delta y_{21} & \Delta y_{22} & \cdots & \Delta y_{2v}\\ \vdots & & & \\ \Delta y_{n1} & \Delta y_{n2} & \cdots & \Delta y_{nv}\end{bmatrix}
$$

建立线性方程:

$$
{}_{n}\hat{Y}_{v}={}_{n}E_{m}B_{v}, \tag{2}
$$

式中

$$
{}_{n}\hat{Y}_{v}=\begin{bmatrix}\Delta\hat{y}_{11} & \Delta\hat{y}_{12} & \cdots & \Delta\hat{y}_{1v}\\ \Delta\hat{y}_{21} & \Delta\hat{y}_{22} & \cdots & \Delta\hat{y}_{2v}\\ \vdots & & & \\ \Delta\hat{y}_{n1} & \Delta\hat{y}_{n2} & \cdots & \Delta\hat{y}_{nv}\end{bmatrix},\quad {}_{m}B_{v}=\begin{bmatrix}b_{11} & b_{12} & \cdots & b_{1v}\\ b_{21} & b_{22} & \cdots & b_{2v}\\ \vdots & & & \\ b_{m1} & b_{m2} & \cdots & b_{mv}\end{bmatrix}
$$

分别表示计算预报量距平矩阵与待定系数矩阵，v 为预报量个数(站数)。(2)式的代数形式为:

$$
\Delta\hat{y}_{kj}=\sum_{i=1}^{m}e_{ki}b_{ij}.\quad k=1,2,\cdots,n\quad j=1,2,\cdots,v \tag{2$'$}
$$

由于 ${}_{n}E_{m}$ 的方差等于特征值

$$
{}_{m}E_{n}E_{m}={}_{m}\Lambda_{m}, \tag{3}
$$

于是用 ${}_mE_n$ 左乘（2）式，得：

$$ {}_mE_n\hat{Y}_v = {}_m\Lambda_m B_v, $$

根据最少二乘法原理，可以证明

$$ {}_mX_n\hat{Y}_v = {}_mX_nY_v, \tag{4} $$

用 ${}_m\Phi'_m$ 左乘上式两边，考虑到

$$ {}_nE'_m = ({}_nX_m\Phi_m)' = {}_m\Phi'_{mn}X'_m = {}_m\Phi'_mX_n = {}_mE_n, $$

得：

$$ {}_mE_n\hat{Y}_v = {}_mE_nY_v, \tag{5} $$

同时，因特征值矩阵具有代数特性：

$$ {}_m\Lambda_m^{-1} = \frac{1}{{}_m\Lambda_m}, \tag{6} $$

这里 ${}_m\Lambda_m^{-1}$ 为 ${}_m\Lambda_m$ 的逆矩阵，于是求得系数矩阵：

$$ {}_mB_v = {}_m\Lambda_m^{-1}E_nY_v. \tag{7} $$

若将 ${}_m\Lambda_m$ 及 ${}_nE_m$ 中的小比重各列略去，则其近似式为：

$$ {}_sB_v = {}_s\Lambda_s^{-1}E_nY_v, \quad (s < m) \tag{8} $$

写成代数式：

$$ b_{pj} = \frac{1}{\lambda_p}\sum_{k=1}^{n} e_{kp}\Delta y_{kj}, \quad p = 1, 2, \cdots, s \quad j = 1, 2, \cdots, v \tag{8$'$} $$

式中 Δy_{kj} 为预报量的观测距平值，为已知，而 λ_p 和 e_{kp} 由正交分解算得，因此 b_{pj} 可求。

由于 ${}_nE_s$ 各列元素具有时间序列的特性，可以采取自回归的方式，在适当选择阶数后，通过一维平稳时间序列法求得 $n+1$ 时间的 e 值：

$$ e_{n+1,p} = \sum_{\alpha=1}^{l_p} a_{\alpha,p}e_{n-l_p+\alpha,p}, \quad p = 1, 2, \cdots, s \tag{9} $$

式中 l_p 为 ${}_nE_s$ 中各列向量的阶数，即预报式的项数，$e_{n-l_p+\alpha,p}$ 为各列向量的近期序列，$a_{\alpha,p}$ 为系数，由相关函数方程组求得。

因此（2）′ 的预报式为：

$$ \begin{cases} \Delta\hat{y}_{n+1,j} = \sum\limits_{p=1}^{s} e_{n+1,p}b_{pj}, \\ \hat{y}_{n+1,j} = \Delta\hat{y}_{n+1,j} + \bar{y}_j, \quad j = 1, 2, \cdots, v \\ \bar{y}_j = \frac{1}{n}\sum\limits_{k=1}^{n} y_{kj}, \end{cases} \tag{10} $$

式中 $\bar{y}_j$ 为各站序列的多年平均值（纵向平均，以下均同）。

由（10）式可知，各站预报式的项数为 $s(s < m)$。这个预报式是作确定月份的逐年预报的，预报期为一年，因此要求序列个数 n 要足够大，才能真正反映出实际的变动规律。如果因子与预报量的数据按横向（时间序列）处理，则作逐月（逐旬）预报。

2. 多元回归因子场的单站预报法

由上所述，自回归因子场的单站预报，必须借助平稳时间序列法建立预报方程，现在试图建立不依赖平稳时间序列法的预报式。

根据(1)和(2)式,有

$$ {}_n\hat{Y}_v = {}_nE_mB_v = {}_nX_m\Phi_mB_v, $$

令

$$ {}_mC_v = {}_m\Phi_mB_v, \tag{11} $$

则

$$ {}_n\hat{Y}_v = {}_nX_mC_v, \tag{12} $$

式中

$$ {}_mC_v = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1v} \\ c_{21} & c_{22} & \cdots & c_{2v} \\ \vdots & & & \\ c_{m1} & c_{m2} & \cdots & c_{mv} \end{bmatrix} $$

也为已知的系数矩阵。以上两式的代数式分别为:

$$ c_{ij} = \sum_{k=1}^{n} \varphi_{ik}b_{kj}, \quad i = 1, 2, \cdots, m \quad j = 1, 2, \cdots, v \tag{11$'$} $$

$$ \Delta\hat{y}_{kj} = \sum_{i=1}^{m} \Delta x_{ki}c_{ij}. \quad k = 1, 2, \cdots, n \quad j = 1, 2, \cdots, v \tag{12$'$} $$

我们的任务不是回报,而是在已知近期因子 $\Delta x_{n+1,i}$ 的条件下预报 $\Delta y_{n+1,j}$。为此,在因子与预报量配对时,务需使因子的时距超前于预报量,各因子超前的时间可以不一,其中超前的最小时距即为预报期。这样,近期因子 $\Delta x_{n+1,i}$ 便是已知的了,由此建立预报式:

$$ \begin{cases} \Delta\hat{y}_{n+1,j} = \sum\limits_{i=1}^{m} \Delta x_{n+1,i}c_{ij}, \\ \hat{y}_{n+1,j} = \Delta\hat{y}_{n+1,j} + \bar{y}_j. \quad j = 1, 2, \cdots, v \end{cases} \tag{13} $$

由于近期因子必须已知,因此这个预报式只能作中短期的预报。

由上式可知,各站预报式的项数仍为 m,而且若对(11)式的系数矩阵 ${}_mC_v$ 进行省略:

$$ {}_m\hat{C}_v = {}_m\Phi_sB_v, \quad s < m $$

并不能减少预报式的项数,反而增加一层误差。

还必须指出,利用(13)式预报的结果,与多元回归的结果是完全一致的。关于这一点,可作如下证明:

由于(12)式可以写成:

$$ {}_n\hat{Y}_v = {}_nX_m\Phi_m\Lambda_m^{-1}\Phi'_mX_nY_v = {}_nX_m(X_nX)_m^{-1}X_nY_v, $$

将上式进行标准化以后(为了书写方便,仍沿用原符号),考虑到

$$ \begin{cases} ({}_mX_nX_m) = {}_mP_m, \\ {}_mX_nY_v = {}_mR_v, \end{cases} \tag{14} $$

式中 ${}_mP_m$ 为标准化因子协方差矩阵,${}_mR_v$ 为因子与预报量的主相关系数。将(14)式代入上式:

$$ {}_n\hat{Y}_v = {}_nX_mP_m^{-1}R_v, $$

根据多元回归消去法原理，可以证明：

$$ {}_mP_m^{-1}R_v = {}_mR_v^{(s)}, \tag{15}$$

这里 ${}_mR_v^{(s)}$ 为标准化方程的回归系数，故

$$ {}_n\hat{Y}_v = {}_nX_m R_v^{(s)}. \tag{16}$$

(16) 式就是多元回归的预报式。

根据 (10) 式与 (13) 式进行预报的结果显然是不同的，从预报实验的结果来看，前者过于保守，后者则波动较大。但它们具有一些共同点：其一，只有因子具有场的形式，而各站的预报量却是彼此无关的。因为不论 ${}_mB_v$ 或 ${}_mC_v$ 中，${}_nY_v$ 并没有进行过场的分解，所以把这两个预报式称为因子场的单站预报。其二，参加场分解的因子的选择具有重要的意义，它是预报效果优劣的关键。所以确定因子的数量及因子的显著性，都是使用该预报法的前提。但是这两种预报法本身并不能解决这个问题。

（二）确定因子的场预报

为了把预报量引进场分解中，可以把因子和预报量看成是一个整体，一起进行正交分解构成合成场。这里提出两种预报方法。

1. 自回归合成场预报法

设有 m 个因子，v 个预报量(站数)，都有 n 个序列，用 ${}_nZ_u$ 表示因子和预报量的距平矩阵，其中 $u = m + v$ 为因子与预报量的总个数。对 ${}_nZ_u$ 进行经验正交分解求得 ${}_u\Lambda_u$，${}_u\Phi_u$ 和 ${}_nE_u$，并将 ${}_nE_u$ 和 ${}_u\Phi_u$ 省略为 ${}_nE_s$ 和 ${}_u\Phi_s$，然后对 ${}_nE_s$ 中的各列使用平稳时间序列法，求得：

$$ e_{n+1,p} = \sum_{\alpha=1}^{l_p} a_{\alpha,p} e_{n-l_p+\alpha,p}, \quad p = 1, 2, \cdots, s \tag{17}$$

最后只对预报量建立预报式：

$$ \begin{cases} \Delta\hat{y}_{n+1,j} = \Delta\hat{z}_{n+1,j+m} = \sum\limits_{p=1}^{s} e_{n+1,p}\varphi_{j+m,p}, \\ \hat{y}_{n+1,j} = \hat{z}_{n+1,j+m} = \Delta\hat{z}_{n+1,j+m} + \bar{z}_{j+m}, \quad j = 1, 2, \cdots, v \\ \bar{z}_{j+m} = \bar{y}_j = \dfrac{1}{n}\sum\limits_{k=1}^{n} y_{kj}. \end{cases} \tag{18}$$

式中 $\bar{z}_{j+m}$ 或 $\bar{y}_j$ 为各站预报量的多年平均值。从方法的功能而言，这个预报法连影响因子也是可以预报的，但这不是我们的目的。

利用这种方法预报，因子对预报量间的时距不必超前，预报期与自回归因子场单站预报相同。

2. 分段合成场预报法

为了使合成场的预报方法不依赖于平稳时间序列法，我们提出一个能相对独立的合成场预报法。同样把因子与预报量看成是一个整体，其距平矩阵为 ${}_nZ_u$，$u = m + v$。通过经验正交分解求得 ${}_u\Lambda_u$，${}_u\Phi_u$，和 ${}_nE_u$，以及省略矩阵 ${}_nE_s$ 和 ${}_u\Phi_s$。这里无疑是 $s \ll u$，但本法的关键是还要求 $s < m$，即省略后的列数 s 还应小于因子的个数 m；而且因子应超前于预报量进行配对，即近期因子 $\Delta x_{n+1,j}$ ($j = 1, 2, \cdots, m$) 应该是已知的。

根据经验正交分解原理，应有

$$ {}_n\hat{Z}_u = {}_nE_{su}\Phi'_s = {}_nE_s\Phi_u. \tag{19} $$

在此基础上，首先建立因子方程：

$$ \Delta x_{n+1,j} = \Delta z_{n+1,j} = \sum_{p=1}^{s} e_{n+1,p}\varphi_{jp}, \quad j = 1, 2, \cdots, m \tag{20} $$

在上式中，左边各值是已知的，右边 φ_{jp} 也已知，而 $e_{n+1,p}$ 为未知。由于 $s < m$，有 m 个方程不难求出 s 个未知数。为此建立（20）式的正规化方程组：

$$ D_{q,s+1} = \sum_{p=1}^{s} e_{n+1,p}D_{qp}, \quad q = 1, 2, \cdots, s \tag{21} $$

式中 D 为协方差：

$$ \begin{cases} D_{qp} = \sum_{j=1}^{m} (\varphi_{jq} - \bar{\varphi}_q)(\varphi_{jp} - \bar{\varphi}_p), \\ D_{q,s+1} = \sum_{j=1}^{m} (\varphi_{jq} - \bar{\varphi}_q)(\Delta z_{n+1,j} - \Delta\bar{z}_{n+1}), \\ \bar{\varphi}_p = \frac{1}{m}\sum_{j=1}^{m} \varphi_{jp}, \qquad p, q = 1, 2, \cdots, s \\ \bar{\varphi}_q = \frac{1}{m}\sum_{j=1}^{m} \varphi_{jq}, \\ \Delta\bar{z}_{n+1} = \frac{1}{m}\sum_{j=1}^{m} \Delta z_{n+1,j}. \end{cases} \tag{22} $$

通过线性求解法，可以求出 $e_{n+1,p}(p = 1, 2, \cdots, s)$。然后建立预报量方程（预报方程）：

$$ \begin{cases} \Delta\hat{y}_{n+1,j} = \Delta\hat{z}_{n+1,j+m} = \sum_{p=1}^{s} e_{n+1,p}\varphi_{p,j+m}, \\ \hat{y}_{n+1,j} = \hat{z}_{n+1,j+m} = \Delta\hat{z}_{n+1,j+m} + \bar{z}_{j+m}. \quad j = 1, 2, \cdots, v \end{cases} \tag{23} $$

从形式上看，（23）与（18）式相似，但预报结果完全不同。这个方法的要领是利用 ${}_nE_u$ 和 ${}_u\Phi_u$ 的省略，将省略后的 ${}_u\Phi_s$ 分成两段，由前一段的已知因子及正规化方程组求出 $e_{n+1,p}$ 后，再建立预报方程的，当 m 越大于 s 时，效果越好。

也必须指出，增大因子个数 m 与省略 ${}_nE_u$ 的列数是有一定矛盾的。因为一方面 ${}_nE_u$ 的列数不能任意压缩，省略后的 ${}_nE_s$ 应当保证有较高的还原率；另一方面，按方法的要求，因子个数不仅不能无限增加，因子的显著性更应慎重考虑。因子组配合适与否，对预报效果的影响十分明显（参看预报误差统计表），必须认真对待。此外，因子选择不合适，s 值小不下去；因子选配多了，干扰就大，预报效果总是较差。因此当 s 接近于 m 时，解的唯一性将受到影响。当 $m < s$ 时，本法无解。由于本法需要已知近期因子，只能做中、短期预报。

以上所述的这两种合成场的预报法，既考虑到外界因子场的影响，又考虑了预报量之间的相互联系，因此是属于预报量的场预报。但这两种预报法由于建立合成场，计算容量相当大，而且当序列较短的情况下，因子与预报量的总个数 u 往往超过序列个数 $n(u>n)$，此时，需要经过时空转换处理。此外，从方法本身的要求来看，因子既要求显著，个数又不

能过多，所以上述两种预报法，也可称为有限显著因子的场预报法。

更重要的是，这两种方法（还包括以上两种单站预报法）都不能自动的挑选因子，要靠人为选配显著因子，或依赖其它方法选配，这就使这几种预报方法的使用受到限制。

（三）优选因子的场预报法

从物理含意组配因子是相关分析的基础，但为了弥补人为选配显著因子所引起的误差，也必须依靠方法本身选出显著因子。为此，我们把经验正交分解法与阶段回归法结合起来，取其之长，建立可以挑选因子的最佳场预报方程。

1. 预报方程的建立

和前面几种方法不同，这里需要将因子序列及预报量序列都实行标准化，为了书写方便，基本上沿用上述的符号。令 x_{kj} 及 y_{kj} 分别表示标准化后的因子及预报量，${}_nX_m$ 为标准化因子矩阵，${}_nY_v$ 为标准化预报量矩阵，m 为因子个数，v 为预报量个数，n 为序列个数（$n > v$）。首先，只对 ${}_nY_v$ 进行正交分解，求得预报量的 ${}_v\Lambda_v$，${}_v\Phi_v$，${}_nE_v$，及其省略矩阵 ${}_nE_s$ 和 ${}_v\Phi_s(s < v)$，其矩阵式及代数式分别为：

$${}_nE_s = {}_nY_v\Phi_s,$$

$$e_{kp} = \sum_{j=1}^{v} y_{kj}\varphi_{jp}, \quad k = 1, 2, \cdots, n \quad p = 1, 2, \cdots, s \tag{24}$$

式中 ${}_nE_s$ 的各列向量使其按比重的大小顺序排列。

其次，把 ${}_nE_s$ 当作新的预报量与各因子求相关，以挑选显著因子。其挑选步骤如下：

先考虑 ${}_nE_s$ 中的第一列（$p = 1$）与各因子的第一级相关

$$R_{i1}^{(1)} = \sum_{k=1}^{n} x_{ki}e_{k1}, \quad i = 1, 2, \cdots, m$$

这里 R 称为权重相关系数。

由上式可求得 m 个 $R_{i1}^{(1)}(i = 1, 2, \cdots, m)$，从中选择绝对值最大者，设为 $\max R_{i1}^{(1)} = R_{l_11}^{(1)}$，则第 l_1 个因子 x_{kl_1} 被选上，建立 e_{k1} 的第一级预报方程：

$$\hat{e}_{k1}^{(1)} = R_{l_11}^{(1)}x_{kl_1}, \quad k = 1, 2, \cdots, n$$

显然 $\hat{e}_{k1}^{(1)}$ 与 e_{k1} 必然存在着误差（第一级误差）：

$$\varepsilon_{k1}^{(1)} = e_{k1} - \hat{e}_{k1} = e_{k1} - R_{l_11}^{(1)}x_{kl_1}, \quad k = 1, 2, \cdots, n$$

于是将 $\varepsilon_{k1}^{(1)}$ 与各因子再求第二级相关：

$$R_{i1}^{(2)} = \sum_{k=1}^{n} x_{ki}\varepsilon_{k1}^{(1)}, \quad i = 1, 2, \cdots, m$$

从 m 个 R 中选大者，设为 $\max R_{i1}^{(2)} = R_{l_21}^{(2)}$，则第 l_2 个因子 x_{kl_2} 被选上，其减小误差的第一级回归方程为：

$$\hat{\varepsilon}_{k1}^{(1)} = R_{l_21}^{(2)}x_{kl_2},$$

第二级误差为：

$$\varepsilon_{k1}^{(2)} = \varepsilon_{k1}^{(1)} - \hat{\varepsilon}_{k1}^{(1)} = e_{k1} - R_{l_11}^{(1)}x_{kl_1} - R_{l_21}^{(2)}x_{kl_2}.$$

如此继续进行下去，至第 α 级时，使 $\varepsilon_{k1}^{(\alpha)}$ 达到足够小时为止，其第 $\alpha - 1$ 级回归方程及第 α 级误差分别为：

$$\hat{\varepsilon}_{k1}^{(\alpha-1)} = R_{l_\alpha 1}^{(\alpha)} x_{kl_\alpha}$$

$$\varepsilon_{k1}^{(\alpha)} = \varepsilon_{k1}^{(\alpha-1)} - \hat{\varepsilon}_{k1}^{(\alpha-1)} = e_{k1} - \sum_{\beta=1}^{\alpha} R_{l_\beta 1}^{(\beta)} x_{kl_\beta},$$

此时,在 α 级回归方程中有 α 个因子被选上。根据阶段回归原理,被选上的因子可能重复选上,此时只要将相同因子合并即可。

令 $$\varepsilon_{k1}^{(\alpha)} = e_{k1} - \hat{e}_{k1},$$

则第 α 级的计算值(第 α 级预报方程)为:

$$\hat{e}_{k1} = \sum_{\beta=1}^{\alpha} R_{l_\beta 1}^{(\beta)} x_{kl_\beta}$$

因实际值 e_{k1} 与计算值 $\hat{e}_{k1}$ 之差为 $\varepsilon_{k1}^{(\alpha)}$, 当 $\varepsilon_{k1}^{(\alpha)} \to 0$ 时, $\hat{e}_{k1} \to e_{k1}$.

对于 ${}_nE_s$ 中其它各列 ($p = 2, 3, \cdots, s$) 也参照以上方法求出 $\hat{e}_{kp}$, 但对不同的列, α 项数不同,记作 α_p, 而且 l_β 代表的因子 x_{kl_β} 也可能不同,记作 $x_{kl_\beta}^{(p)}$, 故第 α 级误差的通式为:

$$\varepsilon_{kp}^{(\alpha)} = \varepsilon_{kp}^{(\alpha-1)} - \hat{\varepsilon}_{kp}^{(\alpha-1)} = e_{kp} - \sum_{\beta=1}^{\alpha_p} R_{l_\beta p}^{(\beta)} x_{kl_\beta}^{(p)}, \quad k = 1, 2, \cdots, n \quad p = 1, 2, \cdots, s \tag{25}$$

式中 $\varepsilon_{kp}^{(0)} = e_{kp}$.

同理,计算值的通式为:

$$\hat{e}_{kp} = \sum_{\beta=1}^{\alpha_p} R_{l_\beta p}^{(\beta)} x_{kl_\beta}^{(p)}, \quad k = 1, 2, \cdots, n \quad p = 1, 2, \cdots, s \tag{26}$$

式中

$$R_{l_\beta p}^{(\beta)} = \sum_{k=1}^{n} x_{kl_\beta}^{(p)} \varepsilon_{kp}^{(\beta-1)}, \quad \beta = 1, 2, \cdots, \alpha_p \tag{27}$$

$$\varepsilon_{kp}^{(\beta-1)} = e_{kp} - \sum_{q=1}^{\beta-1} R_{l_q p}^{(q)} x_{kl_q}^{(p)}. \quad \beta = 1, 2, \cdots, \alpha_p \tag{28}$$

由于 $R_{l_\beta p}^{(\beta)}$ 与时间无关,对于时间 $n+1$, (26) 式可写为:

$$\hat{e}_{n+1,p} = \sum_{\beta=1}^{\alpha_p} R_{l_\beta p}^{(\beta)} x_{n+1,l_\beta}^{(p)}. \quad p = 1, 2, \cdots, s \tag{29}$$

如果因子超前于预报量配对, $x_{n+1,l_\beta}^{(p)}$ 为已知,则 $\hat{e}_{n+1,p}$ 可求。

根据经验正交分解原理,应有

$${}_nY_v = {}_nE_s \Phi'_v, \tag{30}$$

其中 ${}_s\Phi'_v$ 与时间无关,对于时间 $n+1$, 上式应为:

$$y_{n+1,j} = \sum_{p=1}^{s} \hat{e}_{n+1,p} \varphi_{jp}, \quad j = 1, 2, \cdots, v \tag{31}$$

这里 y 是标准化变量,应给予复原,故预报方程应为:

$$\begin{cases} \tilde{y}_{n+1,j} = y_{n+1,j}\sqrt{D_j} + \bar{y}_j, \\ D_j = \sum_{k=1}^{n} (y_{kj} - \bar{y}_j)^2, \quad j = 1, 2, \cdots, v \\ \bar{y}_j = \frac{1}{n} \sum_{k=1}^{n} y_{kj}. \end{cases} \tag{32}$$

这里 D_j 称为预报量的离差平方和。

2. 权重相关系数

将(28)式代入(27)式中得:

$$R_{l_\beta p}^{(\beta)} = \sum_{k=1}^{n} x_{kl_\beta}^{(p)} e_{kp} - \sum_{q=1}^{\beta-1} R_{l_q p}^{(q)} r_{l_\beta l_q}, \tag{33}$$

或者写成:

$$\sum_{k=1}^{n} x_{kl_\beta}^{(p)} e_{kp} = \sum_{q=1}^{\beta} R_{l_q p}^{(q)} r_{l_\beta l_q}, \quad (\text{第}\ \beta\ \text{级}) \tag{34}$$

式中

$$r_{l_\beta l_q} = \sum_{k=1}^{n} x_{kl_\beta}^{(p)} x_{kl_q}^{(p)} \tag{35}$$

为因子 $x_{kl_\beta}^{(p)}$ 与 $x_{kl_q}^{(p)}$ 之间的相关系数,通称辅相关系数。

(34)式说明第 β 级(第 β 次被选上)的因子序列与 e_{kp} 序列的乘积之和,等于具有 β 个不同级数的权重相关系数与辅相关系数的乘积之和。以此类推,便有:

$$\sum_{k=1}^{n} x_{kl_\alpha}^{(p)} e_{kp} = \sum_{q=1}^{a_p} R_{l_q p}^{(q)} r_{l_\alpha l_q}, \quad (\text{第}\ \alpha\ \text{级}) \tag{34$'$}$$

$$\sum_{k=1}^{n} x_{kl_1}^{(p)} e_{kp} = R_{l_1 p}^{(1)}. \quad (\text{第一级}) \tag{34$''$}$$

将(24)式代入(34)″中,求得第一级权重相关系数的表达式:

$$R_{l_1 p}^{(1)} = \sum_{j=1}^{v} r_{l_1 j} \varphi_{jp}, \quad p = 1, 2, \cdots, s \tag{36}$$

式中

$$r_{l_1 j} = \sum_{k=1}^{n} x_{kl_1} y_{kj} \quad j = 1, 2, \cdots, v \tag{37}$$

为因子与预报量之间的相关系数,通称主相关系数。

由于 $R_{l_1 p}^{(1)}$ 是最大值,即第 l_1 个因子 x_{kl_1} 同所有 v 个预报量的主相关系数与地点函数 φ_{jp} 乘积之和为最大,可见 x_{kl_1} 作为最佳因子是对所有预报量而言的,而且 $R_{l_1 p}^{(1)}$ 不是由 $r_{l_1 j}$ 的简单算术求和确定,而是与 φ_{jp} 加权求和确定的,φ_{jp} 具有地理概念,起着加权系数的作用,故称 $R_{l_1 p}^{(1)}$ 为权重相关系数,具有场的概念。

对于第 β 级的权重相关系数,由(33)式可知,它具有较为复杂的内容,随着因子的逐个选入,R 的变动呈衰减形式。图 1 是实际计算的一个典型例子,它同平稳分析中的相关

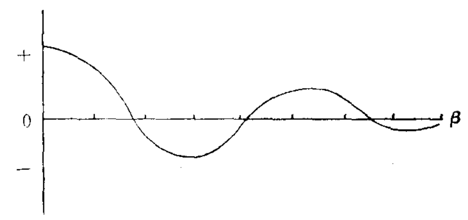

图 1　权重相关系数 R 与选入因子个数 β 的关系曲线

函数与时距的关系极其相似。

3. 误差估计与检定值

由上所述，(25) 式表示第 α 级误差，全系列的误差可以由误差的方差来估计：

$$H_p^2 = \sum_{k=1}^{n} [\varepsilon_{kp}^{(\alpha)}]^2 = \sum_{k=1}^{n} [\varepsilon_{kp}^{(\alpha-1)} - R_{l_\alpha p}^{(\alpha)} x_{kl_\alpha}^{(p)}]^2,$$

考虑到 (27) 式、(28) 式及标准化变量的特性，上式有：

$$H_p^2 = \sum_{k=1}^{n} [\varepsilon_{kp}^{(\alpha-1)}]^2 - [R_{l_\alpha p}^{(\alpha)}]^2$$

$$= \sum_{k=1}^{n} e_{kp}^2 - [R_{l_1 p}^{(1)}]^2 - [R_{l_2 p}^{(2)}]^2 - \cdots - [R_{l_\alpha p}^{(\alpha)}]^2,$$

由 (3) 式可知：

$$\sum_{k=1}^{n} e_{kp}^2 = \lambda_p,$$

上式应为：

$$H_p^2 = \lambda_p - \sum_{\beta=1}^{\alpha_p} [R_{l_\beta p}^{(\beta)}]^2. \quad p = 1, 2, \cdots, s \tag{38}$$

可见 R 愈大，误差愈小，这与一般回归分析的结论是一致的。H_p^2 一般用来作为挑选因子的检定值，但本文在试验中，检定值仍采用均方误的形式，而且不考虑 p 的变动，即

$$G = \sqrt{\frac{H^2}{n}}. \tag{39}$$

同时为了避免由于因子选配不合理，导致 G 值降不下去而循环过多的现象，特采用前后两次均方误之差

$$\Delta G^{(\beta)} = G^{(\beta-1)} - G^{(\beta)} \tag{40}$$

的形式，使 β 增加到满足不等式

$$\Delta G^{(\alpha)} \leqslant a \tag{41}$$

作为挑选因子的检定式。即当满足不等式 (41) 时，再选入因子对减小误差的贡献已不大，可以结束计算。至于检定值 a 应取多少，对不同的预报对象(如温度，盐度等)其值不同，应根据试验确定。

4. 选入因子中显著因子的确定

根据本法的要求，影响因子是分 s 批选入的，每批选入若干个因子。这里的 s 就是 ${}_nE_s$ 中的列数，列数 s 的多少是由误差比

$$W^2 = \frac{\sum_{j=1}^{m} \lambda_j - \sum_{p=1}^{s} \lambda_p}{\sum_{j=1}^{m} \lambda_j} \leqslant 5\% \tag{42}$$

控制的。如果不等式右边的百分比越大，则 s 越小。s 值的大小对预报值是有影响的，在我们水温预报的试验中，$W^2 = 5\%$ 和 $W^2 = 20\%$ 所引起的相对误差可达 0.5℃，一般在 0.5℃ 以下。至于每批选入的因子的个数 α_p 是由 (38) 式或 (41) 式控制的。检定值 a 取

的越大，选入的因子个数就越少。我们曾取 $a=0.001$, 0.005, 0.01 和 0.05 做过试验，选入因子的情况参看表 1。采用 $a=0.001$ 和 $a=0.05$ 所引起的计算误差可达 1℃左右。如果 a 值大了，选入因子个数过少，回报均方误就大；a 值小了，选入因子个数过多，非重要因子的干扰就大，这两种情况都应当避免。关于选入因子重复选上的现象是可能出现的，这是允许的，但在同一列中重复的现象过多，说明 a 值偏小，这一点可作为调整 a 值的参考。根据我们初步试验的结果，在水温预报中，a 取 0.005—0.001 为宜。

在 (38) 式中，由于 $H_p^2 \geqslant 0$，就始终有

$$\lambda_p \geqslant \sum_{\beta=1}^{a_p} [R_{l_\beta p}^{(\beta)}]^2, \quad p = 1, 2, \cdots, s \tag{43}$$

而 λ_p 是按大小顺序排列的，由于经验正交分解法收敛性很好，随着列数 p 的增加，λ_p 作迅速递减，此时上述不等式又始终成立，在 a 值确定的情况下，选入因子的个数必然随着 p 的增加而减少。同时，选入因子的显著性也必然随着 p 的增加而降低。在一般情况下，$p=1$ 时，就有

$$\lambda_1 \Big/ \sum_{j=1}^{m} \lambda_j = 60\%—70\%,$$

因此可以断定，第一批 ($p=1$) 选入的因子是属于显著因子（即重要因子）。为了证实这一点，我们以同样的数据，采用阶段回归法确定的显著因子（在组配的 28 个因子中，显著因子有 13 个，用圆圈起来）与本法的计算结果做一对比（参看表 1）。从表中可以看出，显著因子多落在第一列内，而且集中在 $a=0.005$ 的范围内。

表 1　不同 p 值及 a 值选上因子的分布

	$p=1$	$p=2$	$p=3$	$p=4$
选上因子代号	⑬	9	18	19
	⑭ } $a=0.05$	12 } $a=0.05$	26 } $a=0.05$	15 } $a=0.05$
	②	10	17 } $a=0.01$	6 } $a=0.01$
	③	14	25	7 } $a=0.005$
	⑮	8	2	27
	⑧ } $a=0.01$	15	7 } $a=0.005$	17
	7	⑤ } $a=0.01$	6	18
	㉑	6	1	12
	⑲	10	20	11
	⑳	16 } $a=0.005$	7	8
	⑨	7	6	20
	⑥ } $a=0.005$	5	5	10
	11	16	8	18 } $a=0.001$
	10	14	11 } $a=0.001$	
	25	2		
	20	4 } $a=0.001$		
	④			
	14			
	22 } $a=0.001$			

注：本例中，组配因子 $m=28$，预报量（站数）$v=8$，序列个数 $n=26$。

5. 优选因子场预报方法的特点

这种预报方法的最大特点是：（1）既考虑了预报量（各站）之间的内在联系，又考虑了外界因子的影响，而且对于组配的因子能自行挑选，建立最佳预报方程。由于能自行挑选合适的因子，而那些未被选上的次要因子，对预报方程并无副作用。这样就克服了前面提出的四种预报法对组配的因子务求显著、个数不能过多的不足。即从方法的功能而言，本法不受因子个数的限制，但尽量根据物理含意选配因子仍然是提高预报效果的前提。（2）本法导出的权重相关系数物理概念明确，回报误差估计合乎逻辑，显著因子的确定也很简便，便于推广运用。（3）在使用单站预报法时，往往可能出现少量预报极值，而在使用这个方法进行预报中，预报值基本上不出现极值，可以避免人为的修整，这也是场预报的特点。但必须指出，若在预报量及其近期因子数据中出现错值时，它将扩散到整个场，使场预报结果失真，因此，对预报量及近期因子的整理应格外注意。（4）对于预报量是等级或分类的预报，本法也可使用。因为本法不是将因子与预报量直接求相关，而是将因子与预报量的权重系数求相关，这就避免了因子是数量而预报量是等级的不相称的搭配关系。因此，可将本法运用于水文、气象以及渔获量等的等级预报中去。

二、各种预报方法的预报效果

为了比较以上各种预报方法的实际效果，采用了统一的数据。任取东海区 8 个站（见图 2）1977 年 10 月初的表层水温预报作为典型实例。因子有两组，一组是由阶段回归法对黄海和东海区 80 个站组配的 74 个因子中选取的显著因子，共有 25 个，称为（A）组。这 25 个显著因子是对 80 个站而言的，对我们任取的这 8 个站不一定显著，也不一定合适。另一组因子是为这 8 个站选上的因子，共有 28 个，称为（B）组。这两组的序列都是 26 个。

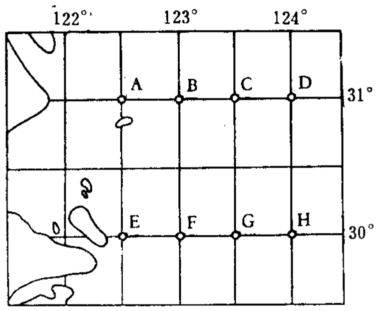

图 2 站位图

（一）预报方法的稳定性比较

以上五种预报方法的原理不同，预报效果必然有别。从预报效果稳定性来看，可将上述方法分为保守型和活跃型两类。为了全面比较，也把不带因子的自回归场预报法及阶段回归法的预报结果一起进行比较。

1. 保守型

从表 2 所列数据可以看出，1977 年 10 月初水温比常年偏低，负距平较大，而所列的三种预报方法的预报值多在平均值附近，即预报距平甚小，属保守型。而且（A）、（B）两组因子的预报结果相差无几，说明保守性大，因子的组配情况对预报结果影响不大。保守性的出现，与组合因子矩阵 ${}_nE_s$ 的平稳性有关；此外，阶数的确定也有影响。一般说来，保守性大的预报方法，其使用价值较低。

2. 活跃型

从表 3 可以看出，所列几种预报方法的预报结果变动性较大，但使用（A）组因子的预

报结果，除优选因子场预报法外，其余预报趋势均相反（正距平），使用（B）组因子的预报效果则较好。可见使用这几种预报方法时，因子的合理组配是提高预报效果的重要课题。优选因子场预报法的优点可以由此体现出来。

表 2　保守型预报距平与实测距平的比较

站号			A	B	C	D	E	F	G	H	$\Sigma/8$	$\|\Sigma\|/8$
实测值			23.5	23.8	23.4	23.6	24.2	24.5	24.8	25.3	24.14	
多年平均值			23.7	24.0	25.0	25.1	24.0	25.5	25.9	26.1	24.91	
实测距平值			−0.2	−0.2	−1.6	−1.5	0.2	−1.0	−1.1	−0.8	−0.77	0.83
预报距平值	不带因子的场预报		0	−0.2	0.2	0.4	0	0.1	0.2	0	0.09	0.14
	自回归因子场单站预报	(A)	0.1	−0.1	0.1	−0.2	−0.7	0.1	0	0.3	0.01	0.14
		(B)	−0.1	0	−0.1	−0.3	−0.1	0	−0.1	0	−0.09	0.09
	自回归合成场的场预报	(A)	0.1	0	0.2	0.2	−0.1	−0.3	−0.1	−0.2	−0.03	0.15
		(B)	0.4	0.5	0.2	0.6	0.5	0.3	0.4	0.3	0.40	0.40

表 3　活跃型预报距平与实测距平的比较

站号			A	B	C	D	E	F	G	H	$\Sigma/8$	$\|\Sigma\|/8$
实测距平值			−0.2	−0.2	−1.6	−1.5	0.2	−1.0	−1.1	−0.8	−0.77	0.83
预报距平值	阶段回归	(A)	0.2	0.9	−0.4	−0.7	0.4	−0.3	1.2	0.3	0.20	0.55
		(B)	1.0	−1.2	−1.1	0.0	0.0	−0.8	−1.2	−1.9	−0.65	0.90
	多元回归因子场单站预报	(A)	2.6	2.9	−0.4	−1.2	1.7	0.9	1.1	2.7	1.20	1.69
		(B)	0.6	0.2	0.7	−0.2	0.5	1.4	−1.0	−1.0	0.15	0.68
	分段合成场的场预报	(A)	−0.3	−0.1	1.2	1.2	0.1	1.4	1.5	1.2	0.78	0.88
		(B)	−0.5	−0.4	−1.7	−1.6	−0.6	−1.7	−1.5	−1.7	−1.09	1.09
	优选因子的场预报	(A)	−1.0	−0.2	−0.7	−0.6	−0.9	−0.7	−0.6	−0.6	−0.66	0.66
		(B)	−0.4	−0.2	−1.2	−1.2	−0.5	−1.6	−1.6	−1.5	−1.03	1.03

（二）各种预报方法的预报误差

将以上各种预报方法的预报误差列于表 4。

由表 4 所列结果可以看出，在多数情况下，使用（B）组因子的预报效果比（A）组的好得多，因为对这 8 个站来说，（B）组的显著因子比（A）组的多一倍。就预报方法而论，在所列方法中，多元回归因子场单站预报法效果最差，因为对任何一站来说，都要考虑所有因子对该站的影响，非显著因子的干扰必然相当大，特别在不合理组配因子(如 A 组)的情况下，误差更大。分段合成场预报法在不合理组配因子（A 组)时，预报误差同样较大，而在合理组配因子（B 组)的情况下，预报效果显著好转。唯有优选因子场预报方法不太过分依赖因子的组配情况，预报效果比较稳定。

表 4 各种预报方法的预报误差统计表(℃)

方法 \ 站号		*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	Σ/8	|Σ|/8
阶段回归	(*A*)	−0.4	−1.1	−1.2	−0.8	−0.2	−0.7	−2.3	−1.1	−0.98	0.98
	(*B*)	−1.2	1.0	−0.5	−1.5	0.2	−0.2	0.1	1.1	−0.13	0.73
不带因子的场预报		−0.2	0.0	−1.8	−1.9	0.2	−1.1	−1.3	−0.8	−0.85	0.91
自回归因子场单站预报	(*A*)	−0.3	−0.1	−1.7	−1.3	0.4	−1.1	−1.1	−1.1	−0.79	0.89
	(*B*)	−0.1	−0.2	−1.5	−1.2	0.3	−1.0	−1.0	−0.8	−0.69	0.76
多元回归因子场单站预报	(*A*)	−2.8	−2.1	−1.2	−0.3	−1.5	−1.9	−2.2	−3.5	−1.94	1.94
	(*B*)	−0.8	−0.4	−2.3	−1.3	−0.3	−2.4	−0.1	0.2	−0.97	0.98
自回归合成场预报	(*A*)	−1.3	−0.2	−1.8	−1.7	0.3	−0.7	−1.0	−0.6	−0.78	0.83
	(*B*)	−0.6	−0.7	−1.8	−2.1	−0.3	−1.3	−1.5	−1.1	−1.18	1.18
分段合成场预报	(*A*)	0.1	−0.1	−2.8	−2.7	0.1	−2.4	−2.6	−2.0	−1.55	1.60
	(*B*)	0.3	0.2	0.1	0.1	0.8	0.7	0.4	0.9	0.44	0.44
优选因子场预报	(*A*)	0.8	0.0	−0.9	−0.9	1.1	−0.3	−0.5	−0.2	−0.11	0.59
	(*B*)	0.2	0.0	−0.4	−0.3	0.7	0.6	0.5	0.7	0.25	0.43

参 考 文 献

[1] 王宗皓、李麦村等，1974。天气预报中的概率统计方法。科学出版社，13—21，68—74，129—149。

[2] Davis, R. E. 1976. Predictability of Sea Surface Temperature and Sea Level Pressure Anomalies over the North Pacific Ocean. *J. Phys. Oceanog.* 6(3): 249—266.

[3] Kutzbach, J. E. 1967. Empirical Eigenvectors of Sea-level Pressure, Surface Temperature and Precipitation Complexes over North America. *J. Appl. Meteor.* 6(5): 791—802.

[4] Демирович, Б. П. и И. А. Марон, 1960. Основы Вычислительной Математики.

A STUDY ON PREDICTION METHODS BY USING THE EMPIRICAL ORTHOGONAL FUNCTION IN CONSIDERATION OF THE INFLUENCE FACTORS

Su Yusong and Su Zhiqing

(Shandong College of Oceanology)

Abstract

The prediction equations formulated in the existing prediction methods by using the empirical orthogonal function are generally based upon field resolution according to the statistical relation among the predictands. The influence of external factors is seldom to be considered. The present paper attempts to introduce the influence factors into the empirical orthogonal prediction method. There are five different approaches to be proposed for this purpose, what the field prediction method with optimum factors is studied with emphasis. In this method we consider not only the internal correlation among the predictands but also the influence of external factors, and the optimal prediction equation which suitable for predicting both values and grades may be formulated by the selected optimum factors from the combinational factors.

The conception of weighting correlation coefficient is suggested here. The obvious difference between it and other correlation coefficients is the former having the field charactor. Several practical examples corresponding to the previous mentioned prediction methods are also presented in this paper.