

一种滑动相关普查自动化统计预报系统的设计

郭天华 陈晓平

(绍兴市气象台)

提 要

本文介绍了一种滑动相关普查方法，该方法可自动排除时段重叠且相关性较差的因子，以及预报系统自动化的实现过程。

一、前 言

随着专业气象服务的发展，长期天气预报已不再限于定时发布几次趋势预报，还要求随服务的需要及时作出用户需要的各种项目预报，这就要求有一个与之相适应的预报手段。因而，我们在研制滑动相关普查自动化统计预报系统过程中，力求系统自动化、客观化和通用性的同时，还要求系统有一定的预报质量。众所周知，统计预报的关键环节之一是因子的预报质量。这里，我们采用滑动相关普查方法，自动排除时段重叠且相关性较差的因子，来采集高质量的预报因子，取得了较好的效果。

二、系统概况

本系统以数据资料库为基础，通过自动滑动相关普查和逐步回归，构成了一个数据管理、方程建立、预报输出等一体化的业务服务自动化流程。全系统共有12个文件，其中批处理文件、自动控制文件各一个，运行文件7个，信息文件3个。系统模型见图1。

三、高质量因子的自动采集

通常人们在做序列型因子的滑动相关普查时，其滑动长度是不可变的，这样，虽然

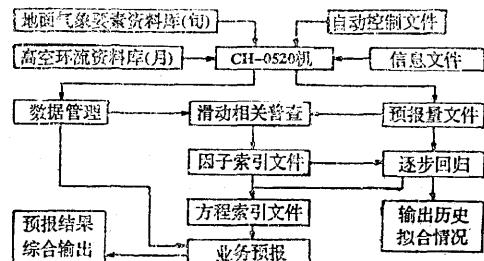


图1 长期天气统计预报自动化系统模型

可以在固定的滑动长度内普查到较好的因子，但是长期天气相关因子中各要素，以及同要素在各季节中的最佳相关时段，可以有很大的差别，而且要定出一个适当的滑动长度往往是不容易的。所以，在本系统中，我们采取了在经验有限区间内，进行可变步长滑动找相关，并自动排除时段重叠且相关性较差的因子，这样既增加了因子的信息量，又可挑选出其中独立性强的高质量因子，供方程选择。

1. 规定滑动步长的有限区间为 k ，则可变的滑动步长有 $1, 2, \dots, k-1, k$ ，共 k 个，资料宽度为 LN (LN 是根据资料内容确定的，月资料 $LN = 12$ ，旬资料 $LN = 36$ ，候资料 $LN = 72$)，则一个预报量在一个因子场中扫描一次可供选的全部因子数为 $k \times LN$ 个（包含当年和隔年）。 N 个因子场的因子信息量为 $N \times k \times LN$ 个，这样，用可变步长滑动，一个预报量的可供选因子信

息量比固定步长滑动时增加了 $N \times (k - 1) \times LN$ 个。

2. 在可选因子量中，先用通过 t 检验的临界相关系数 R_c 为阈值，对因子进行第一次筛选。

假定待查因子与 y 量总体不相关，则它们遵从自由度为 $n - 2$ 的 t 分布*，在指定信度 α 下

$$R_c = \sqrt{\frac{t_{\alpha}^2}{n - 2 + t_{\alpha}^2}} \quad (1)$$

由表查得：当 $\alpha = 0.05$, $n = 30$ 时, $t_{\alpha} = 2.042$
由(1)式求得 $R_c = 0.36$ 。

把 R_c 放入信息文件中，如果因子信息量足够多，则可提高初选阈值水平 R_c 。

3. 因为是可变步长找滑动相关，则不可避免会出现选入的同场因子所在时段重叠。为避免方程所选因子时段集中，独立性差，而使方程稳定性降低，本系统在因子记盘前先进行了排除处理。如果经第一次筛选的同场因子中，经对组合因子的各点进行比较判别，几个因子所在时段有重叠点，则取其中相关系数绝对值最大的因子，剔除其他因子，最后把入选因子再按相关系数绝对值大小排序记盘。

四、系统的自动化

1. 设置信息文件

一个完整的计算系统的完成，常要依赖一系列信息的支持。这些信息又可分为固定信息（例如数组的大小，设备名等）和可变信息，固定信息可以直接固化于程序中，而可变信息的存在往往会影响系统自动化的实现。因此，我们在系统中设置了一系列信息文件（见表1—3）。

将所有的可变信息独立于目标程序，用程序去打开和读信息文件，这样解决了所有

*黄嘉佑编，《气象统计预报试用教材》，北京大学地球物理系。

表1 [MLX]因子场信息文件记录内容

记录号	记 录 内 容
1	因子场数, 西文场名字符数, 中文场名字符数
2, ……, k	因子场西文名, 滑动有限区间, 因子场宽度, 资料存放格式
k+1, ……, N	因子场中文名

表2 [MLY]预报量场信息文件记录内容

记 录 号	记 录 内 容
1	预报量场数, 西文场名字符数, 中文场名字符数, 当年资料截止月份, 相关系数阈值 R_c
2, ……, k	预报量场西文名
k+1, ……, N	预报量场中文名

表3 [MLT]预报季节信息文件

记录号	记 录 内 容
1	预报自然季节, 预报全称

的键盘信息。另外，不论哪个信息有变化，只要编辑信息文件即可，对运行系统毫无影响。

2. 回归方程中因子数的约定及实现

长期预报回归方程中，对20—30年的资料样本，其方程因子数一般以6个为最适宜⁽¹⁾，本系统中，根据各站资料样本的不同，我们约定：

$$M_0 = \begin{cases} INT(M/5) \pm 1 & 4 \leqslant INT(M/5) \leqslant 6 \\ 6 & INT(M/5) > 6 \\ 4 & INT(M/5) < 4 \end{cases}$$

其中， M_0 为因子数； M 为样本长度。

在逐步回归过程中，通常用人为控制 F 值进行多次试验来确定方程的因子数，而本系统中采用了优先法原理，使计算机自动调整 F 值来迎合我们对因子的约定。

假定引进因子的 F 初值为 a ，变化步长

为 b , 引进因子数为 L , 则约定因子数的自动实现过程如图2所示。

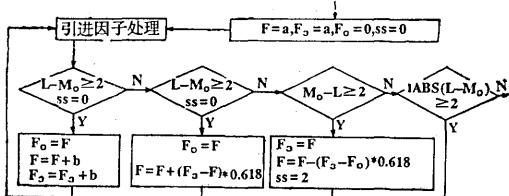


图2 方程约定因子数的自动实现过程

上述运行过程中, 变量 SS 是用来判断引入因子数 L 是否符合约定因子数 M_o 的标准范围, 如果 $L - M_o \geq 2$, 则 $SS = 0$, F 用固定步长 b 增值; 如果 $L - M_o \leq 2$, 则 F 用优先法减值, 且 $SS = 2$, 此后 F 值均用优先法来增值或减值, 直到 $|ABS(L - M_o)| \leq 2$ 为止。

3. 有格式随机文件的使用

由于因子场的信息量远比预报量的信息量大, 而使计算时数据输入所需时间相差很大, 在普查中, 为了节省循环调用资料时间, 我们改变了通常确定一个 Y 量完成所有因子的普查的计算顺序, 而采用先对预报量 Y 循环的方式。但是, 这样一个预报量的因子不同时, 普查完后给记盘带来了困难, 用顺序文件已无法解决。因此, 我们采用了有格式随机文件进行记盘, 这样做不但比顺序文件记盘加快了速度, 并且解决了有格式随机文件在BASIC和FORTRAN语言之间的I/O匹配问题。在本系统中, 回归程序是用FORTRAN语言编写的, 而业务预报程序则采用了BASIC语言(可使屏幕显示美观, 键盘输入和打印输出较灵活), 通过上述方法, 使两种语言之间的信息数据得以正

确的传递。

五、结语

章基嘉等(2)研究提出, 气象要素的长期变化往往是属于平稳的过程, 它的某些统计特征, 至少其预报因子的相对重要性是随时间变化的。这就使得用历史观测资料提炼的统计模式只有较小的时间稳定性。本系统的设计尽管不是在数理统计理论基础上解决这个问题, 但在实际应用时, 可以通过及时更换方程来弥补。因为本系统具有以下特点:

(1) 自动化: 从普查因子到业务预报结果输出, 始终不需人工干预, 且速度大大提高, 解决了方程更新的工作量问题。

(2) 客观化: 本系统输出的是一种客观产品, 无论何人用本系统, 得到的结果是同样的。预报人员用经验分析、使用本系统以提高长期预报的水平。

(3) 实用性: 本系统纯属对历史观测资料和预报量进行气象统计处理, 包含了长期预报中常用的“韵律”和“平稳时间序列”等概念, 系统本身不仅适用于绍兴市, 也适用于其他地市气象台, 只要更换资料即可。

本方法从1986年使用以来, 效果较好, 是我们做长期预报的一种很好的参考工具。

参 考 文 献

- (1) 向元珍, 长期天气预报中逐步回归应用的讨论, 气象, 1986年, 第6期。
- (2) 章基嘉、葛玲编著, 中长期天气预报基础, 气象出版社, 1983年。