

因子的可预报性和预报模型 适用性研究初探

陈孝源

(浙江省气象科学研究所, 杭州 310021)

提 要

本文从因子场与预报量场之间的整体相关性着手, 通过典型相关分析来提取因子信息, 用求得的典型变量作为新预报因子, 经试用表明, 新因子的可预报性比原因子有明显提高。

在预报模型的选择上, 本文提出了依据预报模型对历史样本实际预测精度的优劣来衡量预报模型的预测能力, 从而选用对历史样本预测精度较高的预报模型作未来预报。

关键词: 因子; 可预报性; 预报模型。

一、引 言

在统计预报中, 预报的正确与否, 主要取决于两个方面的因素, 即因子的可预报性和预报模型的适用性。

在因子的选择上, 能常是根据预报因子和预报量之间的单相关系数的大小来确定, 但在实际应用中常会发现, 单相关系数的大小有时往往受个别点的影响, 去掉或增加某个别点, 相关系数就会产生很大变化, 于是包含这些强影响点的数据所导出的统计推断就不够稳定。另一方面, 单相关系数的大小只是考虑了单因子与预报量之间的统计关系, 而忽略了因子之间的共同作用对预报量的贡献。有时某些因子的单相关系数虽然很小, 但它们的共同作用对预报量的贡献却很大。在因子初选时, 由于这些因子与预报量之间的单相关系数过小而往往被抛弃。事实上, 预报量的实测值, 正是各种因素共同作用的结果。为了克服这一缺点, 本文从因子场与预报量场之间的整体相关性着手, 通过典型相关分析^[1,2], 用求得的典型变量作为预报因子, 以期望提高因子的可预报性。

在预报模型的选择上, 怎样在众多的预报模型中选择适当的模型作预报, 这是人们较为关注的一个问题, 但又难以确定, 也就是说没有一个统一的标准。本文从所选用的各种预报模型对历史样本的实际预测能力着手, 根据预测精度的优劣来确定所选用的预报模型。经试用表明, 取得了明显的预报效果。

二、基本资料 and 资料处理

1. 基本资料

预报对象为全省、杭州、嘉兴、宁波、绍兴、温州、丽水、台州、舟山、金华十个行政区域经调和权重法分离后的春粮、早晚稻气象产量, 使用资料为北半球 500hPa 月平均网格点高度场和北太平洋海域 286 格点月平均海表面温度场。供选用的预报模型有, 灰色系统^[3]、车贝雪夫多项式^[4]、预测均方误差 MSEPx 准则^[5,6]、预测平方和 PRESS 准则、逐步回归、混合回归、最优子集回归、岭回归、稳健回归^[7]、非参数回归和动态系统预报的多层递阶法^[8]。

2. 资料处理

首先将十个不同的预报对象分别与北半球 500hPa 月平均网格点高度场和北太平洋海域 286 格点月平均海表面温度场进行单相关分析, 取其单相关系数在 0.4 以上的相关区以其均值作为初选因子, 并通过刀切法对因子的稳定性和预报能力作客观分析, 由此对不同的预报对象选取各自不同的预报因子。然后将预报量和与之对应的预报因子分成南北两片进行典型相关分析。

南片: 温州、金华、丽水、台州。

北片: 全省、杭州、嘉兴、宁波、绍兴、舟山。

典型相关分析是研究两组变量之间相互关系的一种统计方法。它所考虑的着眼点是从因子场与预报量场之间的整体相关性着手来提取因子信息, 通过典型相关分析, 把原来较多的变量归结为少数几个典型变量, 而典型变量组合系数的大小反映了各因子权重的大小。

三、因子的可预报性对比试验

为了进行因子的可预报性比较, 我们用以单相关系数大小所选取的预报因子和经典型相关分析后求得的典型变量分别用逐步回归方法对十个预报区域的春粮、早晚稻的历史样本进行逐年预测。从 N 个样本中除去预测年的样本, 用余下的 $N-1$ 个样本建立回归方程, 就可求得这一年的预测值。

表 1 列出了使用不同因子对春粮 (1965-1989)、早晚稻 (1962-1989) 历史样本进行逐年预测的平均绝对误差。其中 X 为基于单相关系数大小所选取的预报因子, X^* 为经典型相关分析后求得的典型变量 (以下同)。

从表 1 的统计结果可以看出, 把经典型相关分析后求得的典型变量作为新的预报因子, 从对历史样本总体的预测效果而言, 新因子的可预报性比用以单相关系数大小而选取的预报因子有明显提高。

表 1 春粮、早晚稻预测平均绝对误差

预报区域	春 粮		早 稻		晚 稻	
	$\bar{X}$	$\bar{X}^*$	$\bar{X}$	$\bar{X}^*$	$\bar{X}$	$\bar{X}^*$
全省	18.33	11.09	19.45	6.79	32.85	10.84
杭州	23.45	11.76	35.53	14.21	44.06	17.28
嘉兴	32.90	10.34	34.32	15.89	30.80	23.55
宁波	31.29	16.50	33.61	11.27	51.03	16.13
绍兴	29.94	16.28	19.32	9.81	46.28	21.50
温州	13.20	4.60	31.69	13.51	35.10	14.72
金华	25.38	6.45	31.68	13.82	30.92	19.26
丽水	17.76	8.60	24.47	10.72	19.65	12.14
台州	20.21	7.49	35.72	14.29	33.48	20.24
舟山	24.06	21.44	48.58	14.50	66.13	22.29

四、预报模型的预测效果对比试验

由于供选用的预报模型较多,每次作预报时,如若将所有模型都计算一次,不仅计算量较大,而且所求得的预报结果也各不相同,有的甚至得出截然相反的预报结论,这就给实际预报带来一定的困难。在每一个预报模型求得各自不同的预报结果的基础上对各种预报结果进行综合分析后的加权平均,只是反映了对各种预报结果平均状况的不同程度的近似。因此对大丰、大歉年景的预报效果往往不够理想。为此,我们从各种预报模型对历史样本进行逐年预测的平均绝对误差的大小来衡量预报模型的适用性,从而选用预测精度较高的预报模型作预报。

要对历史样本进行逐年预测,计算量很大,我们将预报效果较接近的、选择有代表性的预报模型作逐年预测。从各模型的计算结果可以看出,逐步回归、最优子集回归、岭回归、稳健回归、混合回归和预测平方和 PRESS 准则模型预报效果相近,时间序列方法的灰色系统和车贝雪夫多项式预报效果相近,而多层递阶法预报效果不稳定。经过分析,选定预测均方误差 $MSEP_x$ 准则,逐步回归和非参数回归模型对历史样本进行逐年预测。

表 2 几种模型对春粮的预测效果比较

预报区域	预测均方误差	逐步回归		非参数回归
	$\bar{X}^*$	$\bar{X}$	$\bar{X}^*$	$\bar{X}$
全省	9.76	18.33	11.09	25.32
杭州	9.85	23.45	11.76	27.12
嘉兴	9.88	32.90	10.34	31.50
宁波	15.60	31.29	16.50	34.34
绍兴	11.80	29.94	16.28	27.69
温州	4.62	13.20	4.60	13.18
金华	6.81	25.38	6.45	19.84
丽水	8.22	17.76	8.60	18.14
台州	7.41	20.21	7.49	19.79
舟山	18.78	24.06	21.44	30.23

表 2—表 4 列出了几种模型对历史样本进行逐年预测的平均绝对误差。

从表 2—表 4 可以看出，使用预测均方误差 MSE_{Px} 准则所建立的预报模型从对历史样本逐年预测的平均绝对误差来看其预测精度明显优于其它方法。

表 3 几种模型对早稻的预测效果比较

预报区域	预测均方误差	逐步回归		非参数回归
	X^*	X	X^*	X
全省	6.31	19.45	6.79	29.11
杭州	11.51	35.53	14.21	36.51
嘉兴	14.74	34.32	15.89	34.63
宁波	10.90	33.61	11.27	33.81
绍兴	9.87	19.32	9.81	37.27
温州	11.66	31.69	13.51	35.30
金华	13.09	31.68	13.82	31.78
丽水	10.47	24.47	10.72	23.90
台州	15.86	35.72	14.29	29.52
舟山	12.08	48.58	14.50	35.69

表 4 几种模型对晚稻的预测效果比较

预报区域	预测均方误差	逐步回归		非参数回归
	X^*	X	X^*	X
全省	10.48	32.85	10.84	29.89
杭州	16.89	44.06	17.28	41.49
嘉兴	17.85	30.80	23.55	37.44
宁波	14.07	51.03	16.13	44.12
绍兴	17.25	46.28	21.50	37.27
温州	14.93	35.10	14.72	31.82
金华	13.22	30.92	19.26	24.71
丽水	10.75	19.65	12.14	18.83
台州	11.77	33.48	20.24	30.23
舟山	18.38	66.13	22.29	41.22

五、讨 论

(1) 对不同的预报对象，预报模型的预测效果会不一样，我们曾在新安江流域 5—9 月间的降水预报中，使用混合回归模型^[9]，其预测效果明显优于其它方法。因此，对不同的预报对象，都要从各种预报模型对历史样本的总体预测效果来考虑选用对其为最优的预报模型。

(2) 时间序列方法的灰色系统和车贝雪夫多项式，经使用表明：对呈单调上升或下降的数据序列有较好的预测效果，对数据序列起伏较大，预测效果不佳，这主要反映在对转折点的预测误差较大。

六、结 论

(1) 通过对因子场和预报量场之间的典型相关分析，从因子场和预报量场之间的整

体相关性来提取因子信息,用求得的典型变量作为预报因子,对于提高因子的可预报性具有明显的效果。

(2) 在春粮和早晚稻预报中,从各模型对历史样本总体的预测效果来看,预测均方误差 $MSEP_x$ 准则所建立的预报模型其预测精度明显优于其它方法。特别是对丰歉年景的预测效果尤佳。

参 考 文 献

- [1] 王学仁, 1983, 地质数据的多变量统计分析, 科学出版社, 332-349.
- [2] Glahn, H.R., 1968, Canonical correlation and its relationship to discriminant analysis and multiple regression, *J. Atmos. Sci.*, **25**(1), 23-31.
- [3] 王学明、罗建军, 1986, 灰色系统预测决策建模程序集, 科学出版社, 33-44.
- [4] 中科院大气物理所编, 1987, 李贝雪夫多项式及其在气象水文中的应用文集, 辽宁科学技术出版社, 74-86.
- [5] 陈希甯、王松桂, 1984, 近代实用回归, 广西人民出版社, 106-114.
- [6] 陈孝廉、俞善贤, 1988, 基于预测均方误差 $MSEP_x$ 准则的梅雨期长度统计预报模型, 气象科学研究院刊, **3**, 第 1 期, 100-104.
- [7] 陈希甯、王松桂, 1987, 近代回归分析, 安徽教育出版社, 341-356.
- [8] 韩志刚, 1983, 动态系统预报的一种新方法, 自动化学报, **9**, 第 3 期, 161-168.
- [9] 陈孝廉、沈锦花, 1987, 新安江流域 5-9 月间汛期雨量预报的混合回归模型, 应用概率统计, **3**, 第 4 期, 373-376.

The Research on the Predictability Using Predictors and the Applicability of Forecast Models

Chen Xiaoyuan

(Institute of Meteorological Science of Zhejiang Province, Hangzhou 310021)

Abstract

In this paper, the canonical correlation between the predictand field and predictor field is analysed. We use the canonical variables which have been derived as new predictors. The results of independent sample tests indicate that the scheme can improve the predictability of predictors.

In selecting forecast models, this paper measures the predicting capability of a forecast model according to its real prediction precision of historical samples. Thereby the forecast model with higher precision is used to make future forecast.

Key words: predictor; predictability; forecast model.