

长期预报基本观测资料自动化处理

王建英 王荣英 李翠玉 李月英

(河北省衡水市气象局, 衡水 053000)

摘要 以普遍推广的 A0 文件(即基本观测资料)为基础,经过一定的程序处理,得到各气象要素的气候值,并形成建立方程所需自变量数据文件和预报所需的自变量数据文件。计算自变量与因变量的相关系数,将相关系数较大的因子筛选出来,通过人机交互的方式进行第 2 次筛选,用相关显著的自变量与因变量建立回归方程。由于输出的自变量相对因变量来说都是前期因子,根据回归方程和预报所需的自变量,就可以得出预报结果。

关键词 A0 文件 资料处理 逐步回归 自动化

引言

在以往的工作中,建立一个逐步回归方程需人工处理大量的原始数据,方程的更新也十分烦琐,因此多年使用同一个方程十分普遍。而随着时间的推移,原方程的可用性越来越小,方程亟待更新。针对这种现状,设计了以 A0 文件为基础的长期预报自动生成处理软件(以下简称应用软件)。

应用软件采用 VB 6.0^[1] 语言编写,所采用的技术主要是数据顺序访问技术,逐步回归技术以及点到点的绘图技术等。

利用应用软件所建立的逐步回归方程,是因子经过了双重检验所得到的最优方程。为了避免因子之间因相关程度较高引进后又被剔除,在选择因子时除考虑所选因子与因变量的相关系数显著外,还需考虑各因子之间相关程度小,独立性强,以减少运算量,提升方程的可信度。

目前,国内利用 A0 文件所作的研究仅限于单纯的气候查询,如上海市气象局胡本贵所作的“气象资料处理系统应用软件”,而利用 A0 文件所做的“长期预报自动化处理”应用软件在气象领域尚未见到。

1 程序介绍

1.1 程序设计的整体思路

通过处理 A0 文件,得到建立预报模式所需的前期因子,将这些自变量因子经过两次筛选,与因变

量建立回归方程,流程如图 1 所示。

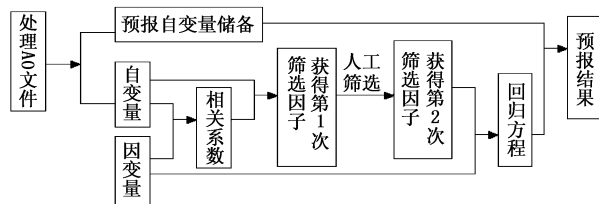


图 1 长期预报自动化处理流程

1.2 资料处理

资料处理是程序的核心内容,即从 A0 文件到生成相关系数自变量数据文件(data1)和预报所需的自变量数据文件(data2)的过程。

(1)在进行资料处理时,考虑了 A0 文件中每个气象要素的不同格式。

(2)所有的因子都经过了同化(即原始数据格式统一化,时段统一化)处理,使同一因子之间的值具有可比性。

(3)处理气候因子充分考虑了缺测资料的影响。

(4)形成的气候因子有:①平均气压,②平均气温,③平均最高气温,④平均最低气温,⑤平均水汽压,⑥平均相对湿度,⑦平均云量,⑧总降水,⑨平均风速,⑩平均 0 cm 地温,⑪平均 5 cm 地温,⑫平均 80 cm 地温,⑬平均日照百分率,⑭最高气温,⑮最低气温。气候值分为月、旬、双旬气候值,弥补了以往气候因子单一的缺点。

(5)资料处理(即自变量处理)的最终结果是输

出 data1 和 data2 两个自变量数据文件。data1 文件中与因变量相对应的每一个自变量均是其前 12 个月的气候因子。自变量的排序不是按时间先后,而是按月份由小到大排列。

自变量的数目是可选的。若在全体自变量处理平台窗体(图 2)中选择“月气候值”,则每月有 15 个气候因子,因子的最大值为 $12 \times 15 = 180$ 个;若选择“月、旬气候值”,则每月有 60 个因子,因子的最大值为 $12 \times 60 = 720$ 个;若选择“月、旬、双旬气候值”,则每月有 90 个因子,因子的最大值为 $12 \times 90 = 1080$ 个。

(6)资料验证。最初,为保证经过程序运算得到的自变量和预报自变量因子百分之百的可靠,增加了“因子质量控制模块”。利用 Excel 处理原始数据方便、可靠的特点,在模块中引入了 Excel。使所检验的因子所对应的 A0 文本文件在 Excel 环境下打开。利用“粘贴函数” f_x 对选定区域的计算功能(SUM, AVERAGE, MAX, MIN 等),便可方便地计算出该因子的气候值并对比检验,当然,缺测和同化也是要考虑的。

利用“因子质量控制模块”对每一类型的因子都进行充分的验证,保证了资料的真实、可靠。

2 运行程序操作步骤

以预报 2005 年冬季即 2005 年 12 月至次年 2 月的平均气温(乘 10 取整)为例介绍运行程序操作步骤。当前系统月期为 2005 年 11 月。

第 1 步:确定图 2 中右边的参数,选择“月,旬气候值”,就决定了因子数最大为 720;选择“因变量的最初月份”主要确定因子名称中的年况(当年、上年、上上年),之后,点击按钮“全体自变量处理”。这一步将生成 data1 数据文件和 data2 数据文件。

第 2 步:点击按钮“进入长期预报平台”。

第 3 步:第 1 次机选因子。首先确定“长期预报平台”窗体(图 3)中左边的参数,其中“样本数”是运行第一步之后自动生成; f_a 是检验因子方差贡献的一个统计量,它依赖于信度 α 和样本数减因子数的差。当 α 为 0.05,样本为几十时,只要因子不是太多, f_a 在 4~5 之间,为方便起见, f_a 一般选择 4.0 作为因子引入或剔除的一个临界值;相关系数临界值一般选择 0.4(因为当样本大于 24 时,相关系数大于等于 0.4 的因子均能通过 0.05 的显著性检



图 2 全体自变量处理平台窗体



图 3 长期预报平台:自变量筛选

验),之后,点击按钮“计算相关系数”。这一步的结果将把相关系数绝对值大于等于 0.4 的因子筛选出来,并输出到右边列表框中。列表框中有因子的序号,因子名称及相关系数(图 3)。

第 4 步:第 2 次以人机交互的方式筛选因子。双击图 3 右边列表框中合适的因子,双击后,被选中的因子进入左边的列表框中。选取的因子参见长期预报平台:方程和预报结果(图 4)左列表框。

第 5 步:建立逐步回归方程。点击图 4“逐步回归”按钮。这一步的结果将把最后所选的因子生成数据文件 data3,并与因变量建立回归方程,输出到右上角的文本框中(替代原右侧列表框的位置,被选入的因子按 data3 中的排列顺序显示)。此外,还有一对比方程(被选入的因子按 data1 中的排列顺

序显示)显示出来,从此方程很直观地看出被选入方程的因子。结果参见图 4 中的“逐步回归方程”和“对比方程”。



图 4 长期预报平台:方程和预报结果

第 6 步:输出预报结果。点击图 4“预报”按钮,这一步是将 data2 中方程所需要的因子代入方程所得的预报结果(参见图 4 右下角“预报结果”)。

3 检验

通过改变运行程序的年、月,可以对前期服从正态分布^[2]的因变量进行预报验证。表 1 达给出 2000~2005 年历年春、夏、秋、冬各季平均气温预报与实况对比。

由表 1 可见,2000~2005 年预报趋势正确百分率为 87.5%(21/24)。

以上结果是在每年每季一个方程(不排除相同季节相同方程)、 $f_c=4.0$ 、相关系数均大于 0.4 的基础上得到的。

表 1 2000~2005 年春、夏、秋、冬各季平均气温预报与实况对比

	2000 年				2001 年				2002 年			
	春	夏	秋	冬	春	夏	秋	冬	春	夏	秋	冬
预报	14.3	26.0	13.0	-0.9	14.1	25.4	13.6	1.2	14.2	26.6	12.5	-1.2
实况	15.5	27.2	13.1	-1.1	15.6	26.5	14.3	1.5	15.6	26.6	12.5	-1.2
趋势判断	√	√	√	√	√	×	√	√	√	√	√	√
	2003 年				2004 年				2005 年			
	春	夏	秋	冬	春	夏	秋	冬	春	夏	秋	冬
预报	14.9	26.9	13.4	0.1	14.7	26.8	13.7	-1.2	14.9	25.5	13.8	-1.2
实况	14.7	25.4	13.2	0.8	14.9	25.1	13.8	-2.1	14.8	27.0	14.7	-1.3
趋势判断	√	×	√	√	√	×	√	√	√	√	√	√

注:趋势判断“√”是指预报和实况相对于其平均值均偏高或均偏低;“×”指预报相对于其平均值偏高(低),而实况偏低(高)。

4 结语

(1)应用软件使建立方程、更新方程以及得到长期预报结果实现了人机交互自动化。

(2)改变运行程序的年、月,应用软件可以对前期服从正态分布的气候因子进行验证。实际工作中,通过验证和总结,找到对预报起关键作用的因子,有时还需要与其他的方法(如方差、谐波、小波分析、均生函数等)配合使用;对不服从正态分布的如月降水量,其立方根大多服从正态分布。

(3)在进行预报时,需要的 A0 文件是连续的、完整的。若上月的 A0 文件暂未审核,需审核之后

方能用。若审核后确认不能用,可修改运行程序的月份,问题就可得到解决。

(4)应用软件尚未剔除因变量的奇异因子所对应的自变量,需要进一步完善。

致谢:感谢谢丽同志在英文翻译中的帮助和指导。

参考文献

[1] 张树兵,戴红,陈哲. Visual Basic 6.0 中文版入门与提高[M]. 北京:清华大学出版社,2001.
 [2] 康锡言,高建华. 山区农业气候区划中年平均气温空间序列的正态性研究[J]. 气象科技,2005. 33(2):185-188.

Automatic Processing of Basic Observation Data Used in Long-Term Forecasts

Wang Jianying Wang Rongying Li Cuiyu Li Yueying

(Hengshui Meteorological Bureau, Hebei Province, Hengshui 053000)

Abstract: By using the specifically processed basic meteorological observation data, the climatic parameters of various meteorological elements are obtained, and the data files of various independent variables needed for establishing the regression equation and forecasting (each independent variable corresponding to a dependent variable is a predictor) are established. The correlation coefficients of independent variables and dependent variable are computed. Through selecting greater correlation coefficient independent variables in a man-machine interactive way, the regression equations are established. According to the regression equation, forecasts can be obtained from the needed independent variables.

Key words: basic meteorological data, data processing, stepwise regression, automation

NASA 卫星首次获得“夜光云”图像

NASA 的 AIM(Aeronomy of Ice in the Mesosphere)卫星于 2007 年 5 月 25 日在 70°N 上首先观测到“夜光云(Night-Shining Clouds)”,地面观测人员则于 6 月 6 日在北欧观测到这些云。AIM 是专门为研究这些特殊云而设计的首颗观测卫星。

这些神秘的云是一种极地中层云(PMCs; Polar Mesospheric Clouds),从空中观测称作“Night-Shining Clouds”,从地面观测则称作“Noctilucent Clouds”。这类云一般形成于北半球夏季地球大气的高层——中间层(Mesosphere),始于 5 月中,持续至 8 月末。此期间,AIM 观测到夜光云的次数越往后越多。在南半球夏季高纬地区也可以观测到这类云。

夜光云如何在极地形成?与从前相比,目前在较低的纬度上也能看到,看到的机会也更多,而且越来越亮,为什么?人们对于夜光云的了解很少。AIM 卫星将在两极地区进行两个完整的夜光云出现期进行观测,计划获取夜光云完整生命期的资料。

AIM 主持人 Hampton 大学的 James Russell III 指出:“很明显,这些云正在变化,而我们不了解它们如何变、为何变、这种变化意味着什么。AIM 资料表明,这种变化与地球低层大气中的变化有关,也许这种变化是地球环境正在发生变化的早期信号。”

AIM 可为科研人员提供关于全球夜光云的云量及其分布特征方面信息,包括构成夜光的微粒的大小与形状。研究人员认为,夜光云最初形成于高纬地区,然后向低纬移动。AIM 项目组人员正在研究夜光云形成与变化的机制及其是否与全球变化有关。2007 年 8 月底北半球的观测任务结束后,AIM 将转向南半球的夜光云观测,从 2007 年 11 月开始,到 2008 年 3 月中旬结束。

曾晓梅编译自 NASA 网站,2007-06-28,

http://www.nasa.gov/mission_pages/aim/multimedia/first_view.html