

段旭, 丁圣, 许美玲. 预报因子选取及方程建立人机交互平台[J]. 气象, 2010, 36(11): 120-125.

预报因子选取及方程建立人机交互平台^{* 1}

段 旭¹ 丁 圣² 许美玲³

1 云南省气象科学研究所, 昆明 650034

2 云南省玉溪市气象局, 玉溪 653100

3 云南省气象台, 昆明 650034

提 要: 利用计算机技术, 将预报因子选取、方程建立和样本拟合融为一体, 建立了预报因子选取及方程建立人机交互平台。该平台包括文件信息, 因子初选和回归分析三个模块: (1) 文件信息模块主要是读取预报对象文件和预报因子文件, 显示并检查文件头信息样本长度的一致性; (2) 因子初选模块主要是计算各气象要素或物理量场中每个格点因子与预报量之间的相关系数, 构成各相关场, 分析这些相关场及其关键区域, 从中筛选出组合因子; (3) 回归分析模块是对选入的组合因子与预报对象进行逐步回归计算, 建立预报方程, 并根据回归方程参数和样本拟合率, 通过改变因子引入和剔除标准来调整回归结果。

关键词: 人机交互平台, 预报因子选取, 预报方程建立

Forecast Factors Selection and the Man-Machine Conversation Platform Establishment by Building Equation

DUAN Xu¹ DING Sheng² XU Meiling³

1 Yunnan Institute of Meteorological Sciences, Kunming 650034

2 Yuxi Meteorological Office of Yunnan Province, Yuxi 653100

3 Yunnan Meteorological Observatory, Kunming 650034

Abstract: Using the calculator technique forecast factor selection, equation establishment and sample fitting into a composite ensemble are integrated, and the forecast factor selection and man-machine conversation platform by building forecast equation are established. This platform has three modules, i. e. document information module, factor primary election module and regression analysis module, which can be described as follows: (1) The document information module is mainly reading the forecast the object document and forecast factor document, examining and displaying the consistency of information length in file header; (2) The factor primary election module is mainly calculating the weather factor or physical quantity field's correlativity between each of grid point factors and the amount of forecast field, and through analysis of this correlation field the combination factors are finally found out; (3) The regression analysis module is carrying on stepwise regression towards forecasting object and combination factor which has picked up from the equation by the second module, and then the forecasting equation is build up, and according to the regression equation's parameter and the rate of sample fitting and through changing the standards for picking up and getting rid of regression results are adjusted.

Key words: man-machine conversation platform, forecasting factors selection, building forecast equation setting

* 云南省科技攻关及高新技术计划 2006SG25 资助
2008 年 3 月 19 日收稿; 2010 年 7 月 15 日收修定稿
第一作者: 段 旭, 主要从事天气预报技术研究. Email: ynmodx@sina.com

引 言

以数值预报为基础的天气预报技术路线,包含两个方面的内容,一是提升数值预报模式的能力,二是提高以优选预报因子为主要手段的数值预报产品解释应用能力。

在数值预报产品的解释应用中,提高能力的最佳途径是选取较好的预报因子。在预报因子初选方法中,目前较多的使用三种方法^[1-8]:一是将格点场气象要素内插到站点上作为预报待选因子;二是将气象要素场的每一格点直接作为预报待选因子;三是利用数据的静态统计方法提取气象要素场主要特征信息。这三种方法通过动力或统计的定量化计算,不同程度改进了预报效果,使数值预报业务模式的产品在预报业务中得到了广泛的应用。但是,在预报因子初选中预报员的经验并没有完全发挥,除了预报因子场可以由主观经验选取外,预报因子场(格点、站点或特征信息)中选取的预报因子则完全由统计方法确定。这在一定程度上削弱了预报员的作用,预报员对天气系统的细节^[9](强度、位置等)和物理意义等主观分析和经验没有完全融合于数值预报产品解释应用中。

为了提高预报员在数值预报产品解释应用中的作用,作者曾于 1996 年提出^[10]了一种简单,但天气、物理意义却比较明确的预报因子选取方法,在多年的预报业务实践^[11]中,证明是有效和可行的。此方法虽然能有效提高预报因子的相关性,但在实际应用中相当繁琐。建立一个预报方程的流程为:编程计算并打印相关场→人工分析相关区并确定关键区→记录关键区相关大值不同组合因子格点坐标→重新编程计算不同组合因子相关系数→人工确定关键区内最优组合因子并记录其入选因子的格点坐标→编程处理从因子场中挑选入选因子→用统计方法建立预报方程。如果要建立县、乡(镇)精细化多要素预报,需要建立几百甚至几千个预报方程,工作量巨大、难以实现。为此,必须利用计算机技术,取代预报因子选取和预报方程建立流程中脱机操作的那一部分繁琐工作,使之成为预报经验与计算机技术结合的预报工具平台。

1 预报因子选择及方程建立

1.1 预报因子选择方法

在因子的选择上,常常是根据预报因子和预报

量之间的单相关系数的绝对值大小来确定。但在实际应用中会发现,单相关系数的大小有时往往受样本资料中个别数据的影响,去掉或增加个别数据,相关系数就会发生很大变化,于是包含这些强影响数据所导出的统计方程就不够稳定。另一方面,单相关系数的大小只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影响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,但它本身及其邻近的因子具有相同的符号,这反映了一种天气背景下的物理意义,显然它们的共同作用对预报量影响很大。在因子出选时,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。为克服这一缺点,可以先计算各气象要素或物理量场中每个格点因子与预报量之间的相关系数,构成各相关场,分析这些相关场及其关键区域,从中筛选出组合因子。

具体做法是:把每一个气象要素或物理量场中每一个格点作为一个预报因子,分别求出每一格点因子与预报量之间的单相关系数,每一格点算完后,就构成了气象要素或物理量与预报量之间的相关场;相关场求出后,选择适当的相关系数筛选因子临界值,把较小的格点去掉;筛选后,依据分析结果,选出天气物理意义明确,成片的正或负相关大值区作为待选的组合因子并计算该区域格点上的气象要素或物理量平均值与预报量之间的相关系数,取绝对值大的作为入选的组合因子。

1.2 预报方程建立

组合因子确定后,利用逐步回归数学模型(也可以用其他统计模型)建立预报方程。在预报方程建立过程中,通过分析引入因子个数、复相关系数、 F -检验值、样本拟合率等信息,调整引入和剔除标准,选择最优方程。

2 平台设计思路

2.1 软件设计原则

给预报员提供发挥预报经验的人机交互式环境,尽可能减少预报因子选取和预报方程建立过程中脱机或烦琐的操作。

2.2 人机交互过程简明

由于平台的功能相对比较单一,各功能模块的

实现设计在同一界面(图 1)上完成,避免人机交互时平台界面换页对工作思路的影响。同时,把每一步过程进行中的信息或中间结果显示出来,供使用者查看或检查。

2.3 尽量体现“场”的概念

预报因子选取人机交互过程是平台设计的核心。预报员在长期的业务过程中形成了对气象要素二维(甚至三维)场的思考方式,因此平台设计时对预报因子与预报对象相关分析的整个过程(格点场计算—格点筛选—组合因子选取)用二维场的形式实现。用“场”的形式进行人机交互,可以使预报员在预报因子选取时与预报经验(各种天气系统在某一区域出现的时间、频率、强度及影响等)有机的结合起来。

3 平台模块功能及操作

平台包括文件信息,因子初选和回归分析三个模块(图 1)。

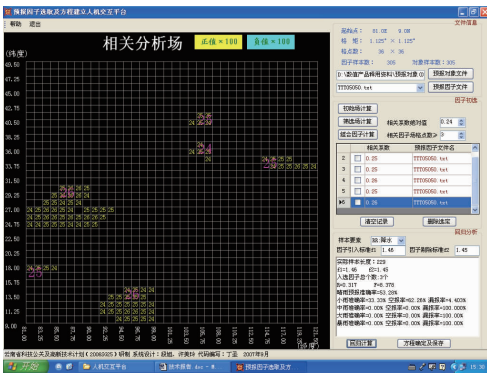


图 1 平台界面

Fig. 1 Platform interface

3.1 文件信息模块

该模块功能包括:(1)读取预报对象文件和预报因子文件;(2)检查对象文件和因子文件头信息样本长度的一致性,并显示读取的文件头信息。

点击“预报对象文件”和“预报因子文件”用于读取预报对象文件和预报因子文件,读取文件的格式正确后文件信息被显示(图 2),其中,预报因子文件可以多选,然后在下拉列表中选择某个预报因子文件。如果文件样本长度和对象文件的样本长度不一致,会给出提示“是否删除此项?”、“文件头格式错

误!”等提示(图略)。

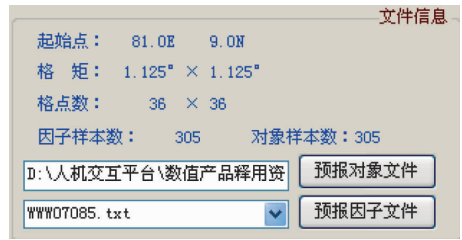


图 2 文件信息模块界面

Fig. 2 The document information module

3.2 因子初选模块

该模块包括:(1)检查预报对象文件和预报因子文件样本长度的一致性;(2)格点初始场相关系数计算并显示;(3)格点筛选场相关系数计算并显示;(4)组合因子相关系数计算并显示;(5)组合因子选取和删除。

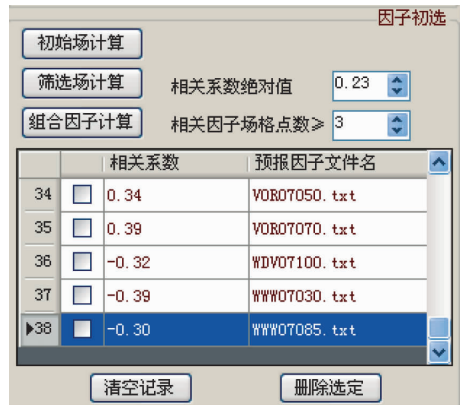


图 3 因子初选模块界面

Fig. 3 The factor primary election module

图 3 给出了因子初选模块界面,点击“初始场计算”框,系统根据因子场文件头信息,计算并确定相关分析显示区网格(图 1 左侧部分),计算并显示初始场格点与预报对象之间的单相关系数(图略)。

在分析单相关系数的基础上,确定筛选相关系数临界值,然后点击“筛选场计算”框,系统根据选定的临界值(如图 3 中的相关系数绝对值 0.23),计算并显示相关系数绝对值大于临界值以上格点的单相关系数(图略)。

确定组合因子格点数(如果格点数为 1,实际上就是单格点因子),计算组合因子相关系数并显示(图 1 左侧部分)。

在相关分析场中点击显示的组合因子相关系数即确定入选因子(图 3),也可以删除或清空已选中

的预报因子。

3.3 回归分析模块

该模块包括:(1)对选入的组合因子与预报对象进行逐步回归计算,建立预报方程,并通过改变因子引入和剔除标准来调整回归结果;(2)计算回归方程的样本拟合率;(3)显示和保存逐步回归分析和样本拟合信息。

组合预报因子选定后,即可操作“回归分析模

块”(图 4)。首先选定预报对象(确定样本要素,目的是为了选择拟合率计算标准),然后确定适当的因子引入标准和剔除标准,最后点击“回归计算”按钮,少时即可显示回归分析信息(包括回归效果参数和回归方程的拟合率)。

如果不满意回归效果,可以改变因子引入和剔除标准来调整回归结果,直到满意为止。

回归方程建立后,点击“方程确定及保存”按钮,自行选取文件名,可保存回归方程及相关信息。

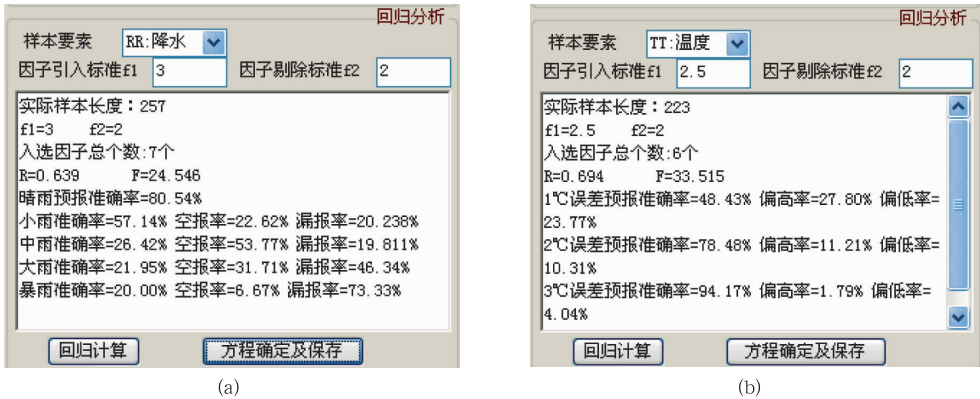


图 4 回归分析模块界面

(a) 降水; (b) 气温

Fig. 4 The regression analysis module for (a) precipitation and (b) air temperature

4 关键技术

该平台使用微软公司的 VisualStudio2005 开发,开发语言使用 C#, VS2005 是个功能强大的 .NET 开发平台,集开发、调试、编译、参考于一体^[12-13]。

文件信息模块主要使用了文本文件读取技术,读取文件头,其思路是用长度为 7 的数组记录文件头中的数字,对于前导空格删去。主要函数:readgenefile(string gfn) 功能是读取预报因子文件并显示,gfn 为预报因子文件路径。函数中使用的变量:string[] headpara = new string^[7] 为预报因子文件头部各参数;string[] objectheadpara = new string^[2] 为预报对象文件头部参数。

预报因子初选技术是最关键的环节,其实现过程为:(1)计算格点相关场并显示;(2)用户根据相关场信息,确定初选因子临界值,计算机自动将低于临界值格点的信息清屏,计算不相邻各剩余格点群组相关系数并自动记录和显示;(3)用户依据组合因

子相关的大小、与预报对象的天气和物理意义,决定取舍,选取的组合因子计算机自动放入临时数据库中并显示。

实现预报因子初选功能,使用了图像处理,数据库操作,递归算法等技术方法,解决预报因子的初始场计算、筛选场计算、组合因子计算以及组合因子初选等目标。具体技术方法是:将预报因子文件中的数字填充进三维数组,预报对象文件中的数字填充进一维数组,计算后的相关系数文件保存进一个二维数组,另用一个 bool 型一维数组记录缺测数据的位置,然后在 picturebox 控件中画出来,根据正负值填充不同颜色。点击组合因子场按钮后,还有组合因子场的相关系数计算,用大数字显示值,单击大数字,就插入数据库中,重复选择只插入一个。用户可以直接看到选择了几个组合因子,也可以删除已选择的因子。用户可以选择不同的预报因子文件,但预报对象文件是固定的,如果重新选择预报对象文件,所选的组合因子全部清空。软件从运行安全考虑,确保计算时不溢出,限制组合因子选取数目不超过 50 个。如果大于 50 个,系统会给出提示。

回归分析模块使用了数据库操作和文本读取技术^[14],其主要方法是读出 selectgene 表中的数据,每一行记录一个组合因子,将各个格点加起来就是一个组合因子,有多少行就有多少因子,然后用 zbhg3 函数进行回归和拟合以及根据需要导出方程结果。主要函数 zbhg3(int n , int k , double[,] x , double $f1$, double $f2$)用于逐步回归^[15]计算和对回归方程进行拟合。其中, n 为因子数, k 为样本数, $x[n, 1, \dots, k]$ 为预报对象, $x[1, \dots, n-1, 1, \dots, k]$ 为预报因子, $f1$ 为因子引入标准, $f2$ 为因子剔除标准。

5 数据格式及运行环境

5.1 数据库说明

forecast.mdb 数据库包含两张表,表 correlation 记录格点位置,用于双击时选取判断格点的位置,表 selectgene 记录选中的格点,用于回归计算,表 selectgene 最多记录 30 行记录,两张表的结构完全一样(表 1)。

表 1 表 correlation 和 selectgene 的表结构说明

Table 1 The table structure of correlation and selectgene

列名	数据类型	备注
dataID	自动编号	主键
genefilename	文本	预报因子文件名
objectfilename	文本	预报对象文件名
相关系数	文本	计算出的相关系数
记录格点	备注	块格点信息
xheight	数字	纵坐标
xwidth	数字	横坐标
预报因子文件名	文本	预报因子文件名

5.2 预报对象文件格式

预报对象文件为文本格式,由文件头和文件内容构成。第一行为文件头,只有一个数字,表示文件内容的样本长度。第二行至末行为文件内容,是一串数据(预报对象),数据之间用空格或回车换行来分割,-9999.9 表示缺测数据。

5.3 预报因子文件格式

预报因子文件为文本格式,由文件头和文件内容构成。第一行为文件头,包含 7 个数字,各数字的意义分别是:起始经度,起始纬度,经度格距,纬度格距,经向格点数,纬向格点数,样本长度。

例如: 94.5 18.0 1.5 1.2 14 13 305。

文件内容由按块排列的规则数据组成,数据块之间由一个空行分割,每一数据块表示一个样本。每一个数据块中的数据按经纬度排列,每一个数据块中的最后一行、第一列数据为起始经、纬度格点数据。

5.4 组合因子选取数目限制

从软件运行安全考虑,确保计算时不溢出,限制组合因子选取数目不超过 50 个。

5.5 逐步回归分析和样本拟合信息保存

信息保存文件为文本格式,文件名自行定义,文件内容为:实际样本长度、因子入选和剔除标准、入选因子总个数、复相关系数和 F -检验值、预报方程、入选因子信息、样本拟合率、剔除因子信息。

5.6 运行环境

平台软件是绿色软件,仅包含“预报因子选取及方程建立人机交互平台.exe”(执行程序)、“forecast.mdb”(系统支持数据库)和“help.chm”(帮助文件)等 3 个文件。在 .NET Framework 2.0 和 MDAC 2.6 以上版本的环境中均可直接运行。

显示器分辨率:1024×768。

平台运行因为涉及的数据量较大,考虑到计算时间和内存空间,建议在 2G 以上 CPU,512M 以上内存的计算机上运行。

6 小 结

(1)“预报因子选取及方程建立人机交互平台”为预报员进行数值预报产品释用提供了一个直观、灵活、便捷的工作平台,它不仅能发挥预报员在预报因子选取中的主观经验,而且充分利用了计算机技术和气象统计方法,在省、市、县各级台站具有较好的推广价值。

(2)文献[16]利用该平台,建立了云南省 125 个气象站的降水、温度客观要素预报方程。通过对比试验表明,绝大多数组合因子都优于单因子,组合因子预报方程的质量普遍比单因子预报方程好。

(3)为了推广应用,平台软件设计为绿色软件,能够在普通计算机环境下运行。对数据的要求也很

灵活,可根据因子场文件头信息,计算不同格距和格点数的相关分析场。组合因子格点数的选取也充分考虑到不同工作的需求,若选 1 个格点,就变成传统的单格点因子。平台除了可供数值预报产品动力统计释用,也可用于短期气候预测分析统计回归。

(4) 平台虽然给预报员提供了一个预报因子选取及方程建立人机交互式良好环境,但因因子选取及方程建立的方法单一,人机交互各环节中也有可能不是最简洁,有待实际预报业务推广应用中不断改进或增加功能。

参考文献

- [1] 刘还珠,赵声蓉,陆志善,等. 国家气象中心气象要素的客观预报——MOS 系统[J]. 应用气象学报,2004,15(2):181-191.
- [2] 孙兰东,张铁军. 甘肃省常规天气要素客观分县预报系统[J]. 干旱气象,2004,22(3):55-58.
- [3] 熊秋芬,顾永刚,王丽. 支持向量机分类方法在天空云量预报中的应用[J]. 气象,2007,33(5):20-26.
- [4] 刘爱鸣,潘宁,邹燕,等. 福建前汛期区域暴雨客观预报模型研究[J]. 应用气象学报,2003,14(4):419-429.
- [5] 陈百炼. 降水温度分县客观预报方法研究[J]. 气象,2003,29(8):48-51.
- [6] 李有宏,韦淑侠,王青川,等. 青海省最高最低气温预报方法研究[J]. 气象,2003,29(7):34-37.
- [7] 苗爱梅,胡永祥,郭玉,等. 以数值预报为基础的省级分县温度预报方法[J]. 气象,1998,24(9):41-46.
- [8] 张建海,王国强. 客观预报中多时刻因子的应用及其效果[J]. 气象,2005,31(5):62-65.
- [9] 薛纪善. 和预报员谈数值预报[J]. 气象,2007,33(8):3-11.
- [10] 段旭. 云南盛夏大雨物理量因子的选取及效果检验[J]. 气象,1996,22(7):30-32.
- [11] 段旭,李英,许美玲,等. 低纬高原地区中尺度天气分析与预报[M]. 北京:气象出版社,2004:154-157.
- [12] Simon Robinson, Ollie Cornes. C# 高级编程[M]. 北京:清华大学出版社,2002:824-826.
- [13] 严蔚敏,吴伟民. 数据结构[M]. 北京:清华大学出版社,2002:54-55.
- [14] 施燕妹,陈培. C# 语言程序设计教程[M]. 北京:水利水电出版社,2004,188-212,219-244.
- [11] 徐士良. 常用算法程序集(c 语言描述)[M]. 北京:清华大学出版社,2002:354-357.
- [16] 许美玲,段旭,丁圣,等. 客观预报方程中因子的选取及应用效果分析[J]. 气象,2009,35(9):112-118.