

阮锡章, 邝美清, 纪嘉雯, 等. 气象因子相关系数矩阵的计算及应用[J]. 气象研究与应用, 2023, 44(3): 34–38.

RUAN Xizhang, KUANG Meiqing, JI Jiawen, et al. Calculation and application of correlation coefficient matrix of meteorological factors [J]. Journal of Meteorological Research and Application, 2023, 44(3): 34–38.

# 气象因子相关系数矩阵的计算及应用

阮锡章<sup>1,2</sup>, 邝美清<sup>2,4</sup>, 纪嘉雯<sup>1,2</sup>, 吴启树<sup>3\*</sup>, 张家斌<sup>2,4</sup>, 朱仕杰<sup>1,2</sup>

(1. 尤溪县气象局, 福建 尤溪 365114; 2. 海峡气象开放实验室, 福建 厦门 361012;

3. 福建省气象台, 福州 350001; 4. 三明市气象局, 福建 三明 365000)

**摘要:** 在建立包含基本气象资料、区域自动站降水资料、高空环流指数和数值预报产品格点资料等多源信息的 Access 数据库基础上, 应用 VB6.0 编程, 通过计算气象因子之间的相关系数矩阵等方法, 研究应用于强降水一致区的划分、建立关键气象因子的预报方程并开展集成预报。结果表明, 在专业气象预报服务中准确率达 75%, 实现从预报因子普查–预报模型建立–气象信息发布的自动化, 提高气象预报科技服务的效率。

**关键词:** Access 数据库; 相关系数矩阵; 强降水一致区; 集成预报; 应用气象

**中图分类号:** P409

**文献标识码:** A

**doi:** 10.19849/j.cnki.CN45-1356/P.2023.3.06

在传统气象统计预报和气候资料统计中, 相关系数的计算是一项基础性并普遍采用的方法, 通常采用相关系数矩阵和主成分分析等方法建立各种预报模型。随着自动站气象资料等非常规资料的融入, 资料录入工作量巨大, 尤其需要大批量资料统计的时候, 计算过程很繁琐, 直接影响到气象研究工作的意愿和效果。近年来, 有关相关系数矩阵算法在人体行为识别与定位深度学习应用方面取得一些进展<sup>[1-2]</sup>, 但在气象因子统计预报等应用方面鲜有报道。本文采用 VB6.0 编程方法<sup>[3-4]</sup>计算相关系数矩阵, 建立多源气象预报服务系统(图 1), 应用于气象因子统计预报服务工作中, 统计结果直观而便于查询。

## 1 资料处理与计算方法

### 1.1 建立资料数据库

(1) 建立气象因子资料数据库。按数据功能分逐日气象资料库、年月旬气象资料库、高空环流指数(74 个特征量)资料库、实时气象资料库和数值天气预报产品格点资料库等, 通过 Microsoft Jet 数据库引擎与数据库链接、应用 DAO 接口编程操作数据

库, 实现 1957—2012 年地面气象观测报表 A0 或 D 文件向逐日气象资料数据库文件、逐日气象资料库向年、月、旬气象资料库之间的自动转换和数值天气预报产品格点资料的读取和保存。

(2) 区域自动站气象资料为三明市气象资料平台输出的 Excel 格式 2010—2019 年逐日雨量。

### 1.2 建立气象因子的文本文件格式

(1) mnkt 文本文件格式( $m$ : 因子数,  $n$ : 样本数,

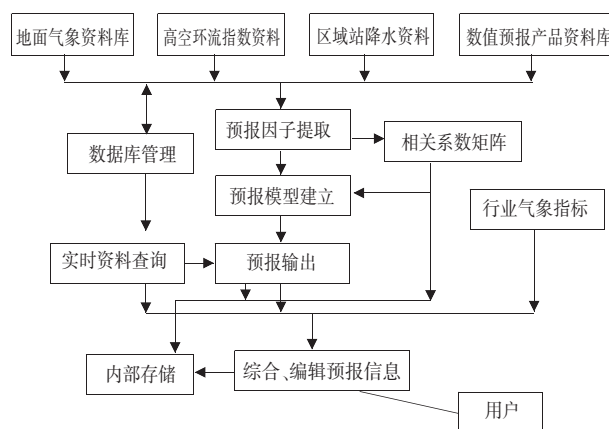


图 1 多源气象预报服务系统流程图

收稿日期: 2023-05-05

基金项目: 海峡气象开放实验室开放课题(项目编号: 2020MXN05)资助

第一作者简介: 阮锡章(1963—), 高级工程师, 主要从事应用气象和天气预报研究。E-mail: rxz0622@163.com

\* 通讯作者: 吴启树(1973—), 正研级高级工程师, 中国气象局首席预报员。

$k$ :开始年代或标记, $t$ :Text,文件说明),用于计算气象因子之间相关系数:

第 0 行: $m \quad n \quad k \quad \text{Text}$

第 1 行: $D_{11} \quad D_{12} \quad D_{13} \cdots D_{1n}$ (气象因子 1 数据)

第 2 行: $D_{21} \quad D_{22} \quad D_{23} \cdots D_{2n}$ (气象因子 2 数据)

.....

第  $m$  行: $D_{m1} \quad D_{m2} \quad D_{m3} \cdots D_{mn}$ (气象因子  $m$  数据)

(2)ynkt 文本文件格式(各字母含义同上, $y=m+1$ ,用于计算预报量与气象因子之间相关系数以及建立回归方程等:

第 0 行: $y \quad n \quad k \quad \text{Text}$

第 1 行: $Y_1 \quad Y_2 \quad Y_3 \cdots Y_n$ (预报量数据)

第 2 行: $D_{11} \quad D_{12} \quad D_{13} \cdots D_{1n}$ (气象因子 1 数据)

第 3 行: $D_{21} \quad D_{22} \quad D_{23} \cdots D_{2n}$ (气象因子 2 数据)

.....

第  $m+1$  行: $D_{m1} \quad D_{m2} \quad D_{m3} \cdots D_{mn}$ (气象因子  $m$  数据)

### 1.3 气象因子相关矩阵计算方法

(1)气象因子的相关系数计算公式如下:

$$r = \frac{\sum_{i=1}^n (x_i - x_p)(y_i - y_p)}{\sqrt{\sum_{i=1}^n (x_i - x_p)^2 \sum_{i=1}^n (y_i - y_p)^2}} \quad (1)$$

( $x_p$  是  $x$  的平均值, $y_p$  是  $y$  的平均值, $n$  是样本数)。

上式的 VB 编程可在有关编程书籍中查询导入<sup>[5-6]</sup>,再用两个循环语句计算出相关系数矩阵,逐步回归等方法亦可同理建立(程序设计流程简图如图 1)。

(2)气象因子格式文本文件 ynkt、mnkt 的导出

主要任务是将预报对象与地面气象资料、高空环流指数或数值预报产品格点资料(选定区域)等进行相关运算,提取通过信度检验、相关较好的预报因子与预报对象数据系列对齐,按相关系数从大到小顺序排列输出,同时输出预报对象和预报因子文件名(含路径及文件类型,下同)及数值预报产品格点资料的经、纬度等资料,保存为气象因子 ynkt 或 mnkt 格式(将 ynkt 文件的第 1 行删除)文本文件。对于未建立数据库的气象因子用 Excel 等相关软件建立相应的格式文本文件<sup>[5-6]</sup>。

(3)采用 ynkt 文件及应用 mnkt 文件建立气象因子相关矩阵文件,通过相关运算,可得出所需的预报方程等资料。同时为方便查询,可将文本文件资料

转换成 Excel 文件。

## 2 应用效果分析

### 2.1 应用于强降水一致区的分析

(1)强降水一致区的确定。本文所称强降水一致区是指受某一天气系统影响时,其降水出现时间和强度都比较相近的区域。研究表明<sup>[7]</sup>,有关区域暴雨影响天气系统及强降水一致区在水利、水电生产有着重要的应用。将各区域站之间暴雨、大暴雨降水量两两相关系数均通过 0.01 信度检验的连片区域确定为优选(或一级)暴雨一致区,未达这一条件但均通过 0.05 信度检验的连片区域确定为普通(或二级)暴雨一致区,进而确定暴雨高度一致区。

(2)采用三明各县市气象局及各区域自动气象站(以下称各区域站)2010—2019 年共 10 a 雨量资料,以县域内任一区域站日雨量  $\geq 50 \text{ mm}$ 、 $\geq 100 \text{ mm}$  确定为该县暴雨、大暴雨日,并进行质量控制,剔除异常值,分析暴雨、大暴雨的地理、时间分布特征,形成 mnkt 格式文本文件。

(3)应用 mnkt 格式文本文件编程计算各县(市、区)域内各区域站暴雨日雨量两两之间相关系数(以下称相关系数矩阵)以及县域之间各区域站暴雨日雨量(经同步处理后)相关系数矩阵(见表 1)。

表中:样本数  $n=117$ ,信度显著临界值  $r_{0.05}=0.197$ , $r_{0.01}=0.256$ 。

(4)综合相关指数(Km12)的计算。由于不同季节暴雨局地差异大,三明市各县(市、区)域之间区域站暴雨日雨量的相关系数,有的出现负数的情况,为了解决此问题,采用公式计算综合相关指数(Km12),进行统一化处理:

$$Km12 = \sum_{j=1}^{m-1} \sum_{i=1}^{m-1} R_{sij} / (m * (m-1)) \quad (2)$$

(2)式中“ $m-1$ ”表示不包括斜对角线上自相关为“1”个例, $R_{sij}$ 以相关系数  $r$  是否通过某一信度检验分别取值如下:

$$f(r) = \begin{cases} 0 \\ 0.4 \\ 0.7 \\ 1 \end{cases} \text{ when } \begin{cases} r < r_{0.05} \\ r_{0.05} \leq r < r_{0.01} \\ r_{0.01} \leq r < 2 \times r_{0.01} \\ r \geq 2 \times r_{0.01} \end{cases} \quad (3)$$

公式(3), $r_{0.05}$ 、 $r_{0.01}$ 分别为相关系数  $r$  通过信度  $\alpha_{0.05}$ 、 $\alpha_{0.01}$  的检验值。令  $R_{sij}=f(r)$ ,计算综合相关指数(表 2)。

表 2 中,以 0.6 为指标,将三明市划分三个强降

表 1 宁化、泰宁县部分区域自动气象站暴雨日雨量相关系数矩阵

台站	济村乡	河龙乡	水茜镇	安乐镇	淮土镇	杨城村	南田村	宁化县	朱口镇	梅口乡	大布村	新桥乡	龙安村	泰宁县
济村乡	1	0.5725	0.4993	0.6849	0.7375	0.64	0.8068	0.7768	0.0982	0.1534	0.1665	0.0812	0.0936	0.141
河龙乡	0.5725	1	0.7586	0.4548	0.4638	0.7065	0.5465	0.5546	0.2264	0.301	0.3122	0.2232	0.3329	0.1952
水茜镇	0.4993	0.7586	1	0.4034	0.4126	0.7174	0.4814	0.5039	0.237	0.2616	0.3385	0.2049	0.3948	0.1927
安乐镇	0.6849	0.4548	0.4034	1	0.8422	0.476	0.7549	0.8253	0.0654	0.0826	0.1287	0.0082	0.1987	0.0636
淮土镇	0.7375	0.4638	0.4126	0.8422	1	0.5235	0.8851	0.8722	0.0654	0.0903	0.0991	0.004	0.0773	0.0796
杨城村	0.64	0.7065	0.7174	0.476	0.5235	1	0.6075	0.5935	0.2187	0.238	0.3065	0.2346	0.3182	0.181
南田村	0.8068	0.5465	0.4814	0.7549	0.8851	0.6075	1	0.8536	0.1035	0.1363	0.1284	0.0344	0.0739	0.0831
宁化县	0.7768	0.5546	0.5039	0.8253	0.8722	0.5935	0.8536	1	0.1274	0.191	0.2168	0.1053	0.1759	0.1515
朱口镇	0.0982	0.2264	0.237	0.0654	0.0654	0.2187	0.1035	0.1274	1	0.5448	0.4535	0.8093	0.2775	0.7206
梅口乡	0.1534	0.301	0.2616	0.0826	0.0903	0.238	0.1363	0.191	0.5448	1	0.7889	0.5008	0.4827	0.8193
大布村	0.1665	0.3122	0.3385	0.1287	0.0991	0.3065	0.1284	0.2168	0.4535	0.7889	1	0.4292	0.7794	0.6826
新桥乡	0.0812	0.2232	0.2049	0.0082	0.004	0.2346	0.0344	0.1053	0.8093	0.5008	0.4292	1	0.3481	0.6857
龙安村	0.0936	0.3329	0.3948	0.1987	0.0773	0.3182	0.0739	0.1759	0.2775	0.4827	0.7794	0.3481	1	0.4195
泰宁县	0.141	0.1952	0.1927	0.0636	0.0796	0.181	0.0831	0.1515	0.7206	0.8193	0.6826	0.6857	0.4195	1

表 2 三明市各县市区暴雨日雨量综合相关指数

县区	尤溪县	永安市	大田县	沙县区	三明市区	泰宁县	清流县	宁化县	明溪县	将乐县	建宁县
尤溪县	0.8650	0.7091	0.7551	0.6254	0.6836	0.0777	0.4328	0.2840	0.4838	0.2425	0.2396
宁化县	0.2344	0.3986	0.1833	0.4444	0.5643	0.1317	0.7267	0.8709	0.6444	0.3133	0.1657
泰宁县	0.0777	0.0925	0.0513	0.0738	0.1203	0.9110	0.0964	0.1317	0.3521	0.7352	0.6774

水一致区:a,东南部包括尤溪县,永安市,大田县,沙县区,三明市区。b,中部区包括清流县,宁化县,明溪县。c,北部区包括将乐县,建宁县,泰宁县。同理,可根据县域内各区域自动站暴雨日雨量的相关系数矩阵及天气系统影响、地形小气候<sup>[8-10]</sup>等情况进一步划区各县域内强降水一致区。

## 2.2 应用于短期气候预测

(1)根据已建立的基本气象资料 Access 格式数据库文件,编程建立气象科技服务支持系统,即通过筛选各气象要素(约 630 多个因子),选出通过 0.05 信度检验的预报因子 21~36 个不等,按相关系数大小顺序,与预报对象对齐输出,形成 ynkt 格式文本文件,保存在指定的预报因子文件夹中。

(2)把 ynkt 文件稍作变换成 mnkt 文件,应用相关系数矩阵程序建立气象因子相关矩阵,以此选出优选因子。

(3)选择以上保存的 ynkt 文件,采用“逐步回归”或“多元-逐步回归综合”(图 2),按提示输入 F 检验值和优选引进方程的因子序号(可反复多

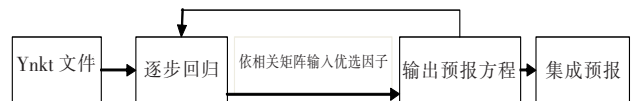


图 2 集成预报流程示意图

次,直到输出的预报方程比较满意为止),输出回归方程。

(4)应用上述方法建立冬季低温冻害预报方程:

$$Y_c = 7.036 - 0.14X_1 - 0.667X_2 - 0.244X_3 \quad (4)$$

(4)式中, $X_1$ 为 2 月日照百分率, $X_2$ 为 4 月雷暴日数, $X_3$ 为 9 月雷暴日数; $Y_c$ 为冬季寒冷度, $Y_c > -2.5$ ,则没有冬寒, $Y_c < -6.5$ ,则有 2 级冬寒;方程复相关系数  $R_f = 0.7463$ ,经 F 检验达到极显著水平(信度 0.001)。

近四年试用效果(表 3),2019、2020、2022 年预报值与实况相符,2021 年预报与实况不符,但预报值比前一年高 1.6℃,结合尤溪县金柑等果树冻害风险分析<sup>[11]</sup>,为果树安全越冬提供气象依据,实际服务效果良好。

表 3 冬季低温预报方程应用检验

年份	$X_1$	$X_2$	$X_3$	$Y_c$	实况
2019	13	9	4	-1.763	-1.3
2020	37	7	12	-5.741	-4
2021	48	3	10	-4.125	1.5
2022	16	11	0	-2.541	-2.4

3 结论与讨论

(1)采用 VB6.0 计算机编程技术,应用 Access 数据库、Excel 和文本格式等气象因子资料,实现气象因子普选及相关矩阵的计算自动化。

(2)应用气象因子普选和相关矩阵计算资料,综合采用多元与逐步回归分析、集成预报等方法,可以应用于区域站强降水一致区的划分、短期气候预测分析及在应用气象服务资料的查询和统计,可建立大量气象统计预报方程,实现从预报因子普查到气象信息发布的自动化,有利于提高气象服务社会经济效益。

在当前气象自动站大量增加,气象区域站资料得以数十倍增加的情况下,尤其需要通过编程相应软件来完成资料的统计分析,本文仅是初步尝试,开展数值预报产品的释用和 OTS 资料应用等方面有待进一步探讨。

参考文献:

[ 1 ] 蔡巍巍. 基于深度学习的人体行为识别与定位方法[J].

工业控制计算机, 2023,36(1):100-101,104.

[ 2 ] 梁鑫宇,宋明丹,韩梅,等.不同农作物秸秆腐解规律及模型预测评价[J].核农学报,2023,37(6):1244-1252.

[ 3 ] 木林森,罗丽琼,奚红宇,等.Visual Basic6.0 入门与技巧[M].北京:清华大学出版社,1999:92-150.

[ 4 ] 刘炳文,许蔓舒. Visual Basic 程序设计教程[M]. 北京:清华大学出版社,2000.

[ 5 ] 高春艳,李俊民,刘彬彬,等. Visual Basic 程序开发范例宝典[M].北京:人民邮电出版社,2006.

[ 6 ] 李鸿吉.Visual Basic6.0 数理统计实用算法[M].北京:科学出版社,2006.

[ 7 ] 吴滨,文明章,李玲,等. 福建省不同短历时暴雨时空分布特征[J]. 暴雨灾害, 2015,34(2):153-159.

[ 8 ] 吴启树,郑颖青,沈新勇,等.福建一次秋季大范围暴雨成因分析[J].气象科学,2010,30(1):126-131.

[ 9 ] 李玲,文明章,张容焱,等.闽南沿海暴雨特征及可能最大降水估算[J].气象科技,2011,39(1):13-18.

[10] 廖廓,何金海.福建秋冬之交降水强度变化及其环流背景分析[J].气象科技,2017,45(5):864-860.

[11] 阮锡章,洪维群,郑文君,等.福建省尤溪县金柑冻害风险分析与区划[J].中国农业资源与区划,2017,38(11):165-170.

## Calculation and application of correlation coefficient matrix of meteorological factors

RUAN Xizhang<sup>1,4</sup>, KUANG Meiqing<sup>2,4</sup>, JI Jiawen<sup>1,4</sup>, WU Qishu<sup>3</sup>, ZHANG Jiabin<sup>2</sup>, ZHU Shijie<sup>1,4</sup>

(1. Youxi Meteorological Bureau, Fujian Youxi 365114, China; 2. Strait Meteorological Open Laboratory, Fujian Xiamen 361012, China; 3. Fujian Meteorological Observatory, Fuzhou 350001, China; 4. Sanming Meteorological Bureau, Fujian Sanming, 365000, China)

**Abstract:** Using VB6.0 programming, an Access database containing multi-dimensional information such as basic meteorological data, regional automatic station precipitation data, high-altitude circulation index, and numerical prediction product grid data was established. By calculating the correlation coefficient matrix between meteorological factors and other methods, it was applied to the division of consistent areas of heavy precipitation, establishing prediction equations for key meteorological factors, and conducting integrated forecasting. The results show that the accuracy rate of professional meteorological forecasting service reaches 75%, achieving the automation from the census of forecasting factors, the establishment of forecasting models to the release of meteorological information, and improving the efficiency of meteorological forecasting technology services.

**Key words:** access database, correlation matrix, consistent area of heavy precipitation, integrated forecasting, applied meteorology