

文章编号: 2096-1618(2023)03-0324-06

成都区域气象站自动分型方法设计及应用

夏昕¹, 王冬萌², 贺南³

(1. 成都市气象局, 四川 成都 611130; 2. 成都信息工程大学通信工程学院, 四川 成都 610225; 3. 成都温江国家观象台, 四川 成都 610225)

摘要:为快速区分下垫面不同气象站类型,采用一种区域气象站自动分型方法。选取两个日气温极小值之差在 $-5\text{ }^{\circ}\text{C}\sim 5\text{ }^{\circ}\text{C}$ 站点,以 $0.1\text{ }^{\circ}\text{C}$ 为步长划分档位,按照温差值的高低顺序排列温差;统计90天内,各档温差值出现的频数,得到反映某个温差位拥有多少样本量的频次分析序列,取样本量的中位数频次进行归一化处理,即为特征频率,记为 F ;设置量化因子 K_f ,量化距离影响因素,下垫面一致性的影响可用 K_f 体现。最终结果表明,经运算分型站点被自动分为A、B、C 3个大群,通过观察A群和C群可知,乡村型站点与城镇型站点在地理分布上被清晰地自动区分,通过对A群C群进行抽样检查,及对城市站点与乡村站点的数据比对,验证了运算分型的有效性。

关键词:分型;日最低气温差值;频率;下垫面类型

中图分类号:TP311.11

文献标志码:A

doi:10.16836/j.cnki.jcuit.2023.03.012

0 引言

为提供精细化气象服务,提升气象减灾防灾能力,中国陆续建设了大量区域气象自动站,其中成都地区自动气象站的总量已超过500个。为满足社会对气象服务工作提出的更高要求,气象监测网格将进一步加密,台站数量还将继续增长,台站管理和数据应用面临全新压力;尽管区域站连续监测的雨量、风速等实时数据已在防灾减灾等气象服务工作中发挥了重要作用,但由于气温测量数据受局部环境条件的影响很大^[1-2],同时数量巨大的区域自动站并不适宜沿用国家级站点(大监站)的建设和维护标准,区域自动站监测数据的可比性、代表性和数据序列的连续性受到不同程度局限,使区域自动站的数据在精细化预报与服务等领域的应用价值未能充分体现;尤其设置在市区内的站点受场地限制,因道路、公园水体等外部环境受到的影响显著^[3-10],自动站测量值的精细化应用更需要准确匹配站点的背景环境条件,但这对环境变动频繁的区域站而言有一定难度。

在站点数量迅速增长和城市建设发展的背景下,如何从大量建成站点中快速筛选出观测环境已发生较大变化、观测数据质量下降的站点,为站点迁址、优化布局、数据精细化管理等提供参考依据方面进行了研究。通过气象大数据的聚类分析有益于实现站点自动化分型^[11-12],但气象数据的聚类还需注重物理机理^[13]以便于应用解析。贺南等^[14]在对成都地区气温

极小值的站间空间差值序列的频数分析中注意到,较高的频率与两个站观测条件的一致性有关联,并同阴晴等天气条件也有关联^[15-19],频率与站间距离等客观因素也联系密切,因此可考虑使用分析温差的频率特点这种技术线路,实现区域站点依照观测环境条件聚类型,以解决区域自动站数据应用中的实际环境背景的主动识别问题^[20]。

1 区域气象站自动分型方法

1.1 资料

数据资料来源于成都地区14个国家气象站近10年逐日历史资料和252个建站时间较长区域自动站近5年逐小时资料,区域站小时数据剔除异常值后计算得到逐日最低气温。气象站自动分型方法中采用聚类方法对分型量化因子做进一步处理,从而实现聚类型。

1.2 数据处理

分型运算使用的数据为通过两个站点日最低气温的差值生成的温差位/频率序列 $\{X_i, f_i\}$,整理时将站A的日气温极小值与站B的日气温极小值一一对应求差,该差值记为 ΔT_{\min} ,此时 $\Delta T_{\min} \in \{\dots, -0.2\text{ }^{\circ}\text{C}, -0.1\text{ }^{\circ}\text{C}, 0\text{ }^{\circ}\text{C}, +0.1\text{ }^{\circ}\text{C}, +0.2\text{ }^{\circ}\text{C}, \dots\}$ 。通常 $\Delta T_{\min} < 5\text{ }^{\circ}\text{C}$ 且 $\Delta T_{\min} > -5\text{ }^{\circ}\text{C}$,在这个值域内,用 $0.1\text{ }^{\circ}\text{C}$ 为步长,按照温差值的高低,顺序排列温差,温差的档位值记为“ X_i ”,令 $X_i = 0.1i\text{ }^{\circ}\text{C}$, $i = \dots, -4, -3, -2, -1, 0, +1, +2, +3, +4, \dots$ 。单个由两个站点生成的 ΔT_{\min} 会对位

某个档位的 X_i 值一次,即 $\Delta T_{\min} = X_i$ 时,计数 $P_i = 1$;取一个时段共 M 天,将可以用 M 个 ΔT_{\min} 构成 ΔT_{\min} 的时间序列 $\{\Delta T_{\min}^j\}$, ($j=1,2,3,\cdots,M$),在这个序列中,统计出各档 X_i 温差值上 ΔT_{\min} 出现的频数 $f(x_i)$:

$$f(x_i) = \sum_{j=1}^M P_i \Delta T_{\min}^j = X_i \text{ 则 } P_i = 1; \Delta T_{\min}^j \neq X_i \text{ 则 } P_i = 0$$
即可以得到反映某个温差位 X_i 拥有多少样本量的频次分析序列 $\{X_i, f(x_i)\}$ 。观察不同长度的时间序列时,需将频次量转化为归一化的频率值 F ,数据处理流程见图 1。

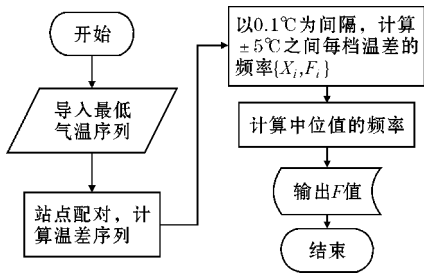


图 1 F 值计算流程

1.3 客观基础

分析所选用的特征频率反映一段时间内两个站点日气温极小值的差值的分布特征,差值集中时特征频率较高,差值分散时特征频率较低,由于以往对这类频次/温差位数据序列进行系统分析的文献不够丰富,因此对于分型方法的客观依据的阐述,以常识性规律的归纳为主,包括:

(1)如果一个站点环境干燥,另一个站点环境湿润,它们之间的气温差值分布相较于两个湿润的站点间的气温差值分布要分散一些;两个荒漠环境下干燥站点间的气温差值分布比两个临海环境下湿润的站点间的气温差值分布分散,这说明地面观测站点的环境条件,与两个站点的气温空间差值的分散程度存在关联,进而也就与特征频率有关联。

(2)即使下垫面条件相近,相对干燥的冬春季,与水汽充分的夏秋季相比较,站点之间产生的气温差值也会相对分散,再次说明观测环境中的含水量等因素与特征频率有关联。

(3)受地形等地理条件影响,对天气系统能同步响应的站点之间产生的温差值会相对集中一些,说明两个站点的地形系数差异对特征频率会有影响。

(4)在众多无风的安静夜晚,空气团的热交换形式主要是长波辐射散热,日气温极小值形成的机理较单一,便于更稳定地提取下垫面的特征信息,同时,日气温极小值对环境敏感,如通常城市中测得的日气温极小值会高于郊外,因此气温极小值为基础的数据序列中隐含有可以用于提取分析观测环境的信息。

(5)天气背景条件为阴天时,站间的日最低气温差值更为集中。使用阴天较多的成都地区国家站30 a的74710组有日照背景的站间最低气温差值样本,同45828组两站均无日照条件的站间最低气温差值样本相比较,后者的温差值分布更集中(参考频次中位数对应的温差位的绝对离差值小30%),这显示出站间日最低气温差值的分布方式,以及特征频率还会附带有天气背景条件的波动量。但特征频率中的环境影响量却是比较稳定的值,因而可以通过对较长时间序列的观察削弱天气背景影响,突出特征频率中的环境影响量。

(6)间距小的站点间,温差值的分布要集中很多,距离对特征频率的影响强势并且恒定,因此,要突出观测环境量的影响必须消除距离影响因素,这是方法的运算基础。

1.4 实验设计

成都地区气候平和,有很多无风和寡照的天气,有较多日最低气温是在绝热环境下的昼夜日周期背景下形成^[15],站间的气温差值分布集中并且比较稳定,有利于提取下垫面的差异信息。当使用一定时间长度序列,平均化天气条件对特征频率 F 的影响使其稳定,再设置量化因子 K_f ,量化距离影响因素 $K_f = \partial F / \partial D$ (式中 F 指特征频率, D 指站间距离),进而下垫面一致性的影响就可以用 K_f 体现。通过选用观测条件一致的站点组合的 K_f 值作为判断两个站下垫面一致性的参照指标,再对本地国家站、区域站间的大量温差数据组合的排序计算,就可以实现区域自动站以足够的一致性指标 K_f 值聚类类型。

通过统计成都地区站点间组合的距离 D 与特征频率 F ,可以粗略得到 F 随 D 的渐变关系 $F = K \times D + 11.3$,如图 2 所示。

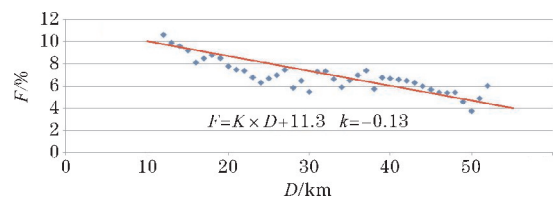


图 2 成都地区国家站呈现的频率与距离的关系

图 2 中,如果 F 值在距离增加时急速衰减,说明两个站点不容易重叠出现固定气温差异值,联系不稳定,即关系式中 K 对一致性有所反映。但斜率 k 不方便观察, K_f 将斜率 k 的变动量转化为便于比对的指标化倍率值,整理成都本地样本的实际数据可以得到 $K_f^2 = ((K \times D)^2 + 23 \times K \times D + 128) \times D$ 。由此这些样本的 K_f 提取式为

$$\begin{cases} K_f = ((K \times D)^2 + 23K \times D + 128) \times D)^{1/2} & (12 \text{ km} < D < 55 \text{ km}) \\ K = \frac{F - 11.3}{D} & (F < 11.3) \end{cases}$$

指标值 K_f 中和了距离权重后,不再随距离变化,图3为成都地区国家站间形成的 K_f 值。

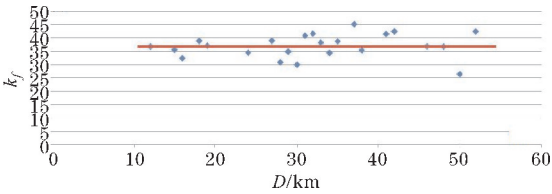


图3 国家站之间呈现的 K_f 指标与距离的关系

K_f 指标量不随距离变化的特点,为观察特征频率中的其他影响因素提供了基础,图4中的A,B,C3个站点,A,B站点为环境理想的站点(国家站),C为观测环境遭受干扰的站点(国家站),在同一时段,A,B站点间的 K_f 值达到45,但A,C之间 K_f 值为35,B,C之间 K_f 值只有32,这组样本中,C站的观测环境异样对 K_f 的影响很明显。



图4 3个样本站的环境影像(1:2256)

K_f 指标整合了 F 值与站间距离的关系,同一间距上 F 越高两个站点的环境相似度会越高;不同间距则 K_f 越大,两个站点的环境相似度会越高,因此 K_f 可用于设置站点间的比对阈值^[21]。成都地区国家站间的 K_f 值主要在30~48,区域站与国家站间 K_f 值分布在8~48。以国家站间的下限值30为参考阈值,聚类分型的实现的方法流程如图4。主要流程步骤包括:基础数

据导入与 $\{X_i, f_i\}$ 序列生成;站点地理信息导入与 K_f 计算; K_f 排序; K_f 值聚类分型及后续分型特征解析。

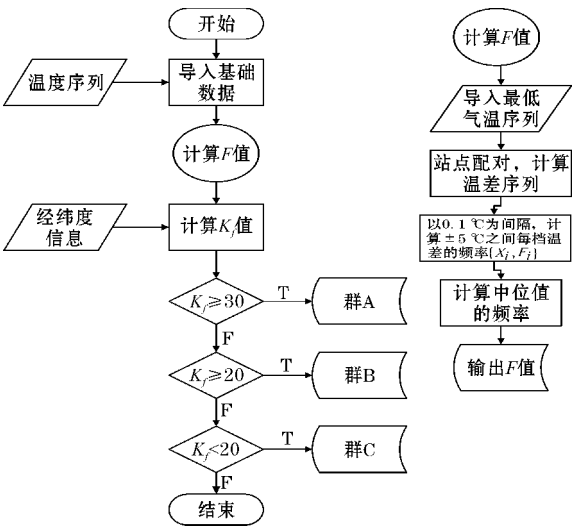


图5 分型计算流程

2 结论与讨论

使用成都地区14个大监站(国家站)和252个区域站的大样本验证,用大监站间的 K_f 为参考标准,经运算分型,站点被自动分为3个大群。第一大群(A群)包含全部大监站和125个区域站,该群观测环境条件的主要特点是乡村站点,农田下垫面为主,市区内没有站点入群,地理位置分布见图6(a);第二大群(B群)包含79个区域站,该群观测环境条件的主要特点是较大面积的公园绿地、山区森林等,自然下垫面为主,地理位置分布见图6(b);第三大群(C群)包含48个区域站点并形成6个子群,其中站点数量最多的是32个城镇站点,地理位置分布见图6(c)、图6(d)。

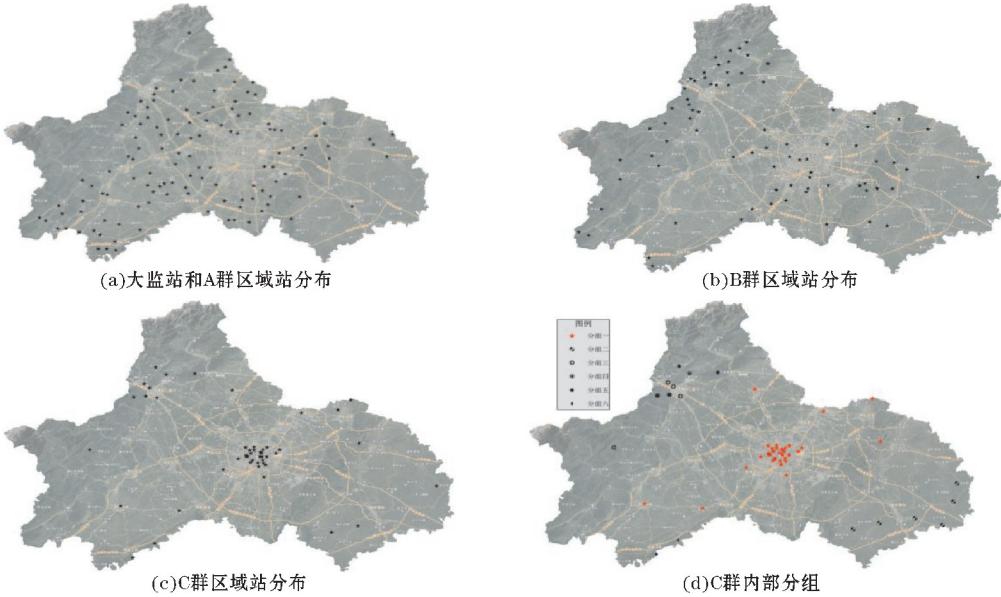


图6 分群站点位置分布

从图 6(a)、图 6(c) 对比看出,通过聚类分型,城镇型站点与乡村型站点在地理分布上被清晰地自动区分,同时这两类站点在观测数据上也有显著差异:图 7 对比了所有站点在 2008 年 2~4 月的平均相对湿度和

平均最低温度,可见 A 群相对湿度明显高于 C 群,A 群最低气温明显低于 C 群。

表 1~4 对 A 群和 C 群数据进行了抽样检查。

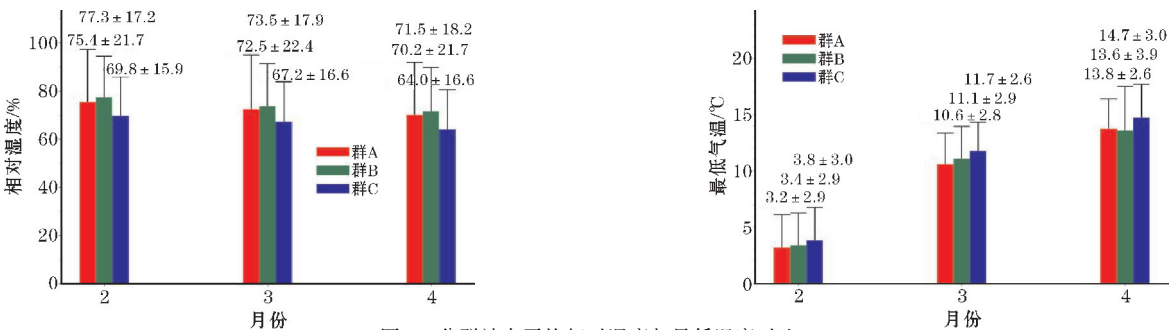


图 7 分群站点平均相对湿度与最低温度对比

表 1 2008 年 2~4 月阶段平均值数据对比

所属分群	站点号	最高气温/℃	最低气温/℃	最高温时段湿度/%	最低温时段湿度/%	日较差/℃
A	SI018	17.4	8.6	53	89	8.8
A	SI024	16.8	8.5	59	98	8.3
A	SI030	16.8	8.3	54	89	8.5
C	SI006	17.8	11.1	48	77	6.7
C	SI007	17.5	9.5	49	83	8.0
C	SI009	17.4	10.9	49	77	6.5

表 1 中的数据反映本地样本在季均值方面,C 群(城镇站)的最高气温略高于 A 群(乡村站)点,最低气温明显高于 A 群;C 群的相对湿度明显低于 A 群,同

时日较差也低于 A 群。说明分型算法区分出的这两类站点,在气温、湿度数据上有明显的整体差异。

表 2 样本 90 d 数据里日气温极大值时段城市站点与乡村站点相对湿度的统计比较

	样本 1	样本 2	样本 3	样本 4	样本 5	样本 6	样本 7	样本 8	样本 9
乡村型站 A	SI018	SI024	SI030	SI018	SI024	SI030	SI018	SI024	SI030
城镇型站 C	SI006	SI006	SI006	SI007	SI007	SI007	SI009	SI009	SI009
湿度 A>C 天数/d	68	75	64	70	73	62	60	71	61
湿度 A<C 天数/d	14	13	24	16	14	22	23	19	27
A>C 的占比/%	76	83	71	78	81	69	67	79	68

统计在表 2 中的数据反映本地样本最高气温时段的相对湿度逐日对比结果,城市站点多数低于乡村站

点。说明分型算法区分出的乡村站点的湿度明显整体高于城市站点,分型运算有效。

表 3 样本 90 d 数据里日气温极小值时段城市站点与乡村站点相对湿度的统计比较

	样本 1	样本 2	样本 3	样本 4	样本 5	样本 6	样本 7	样本 8	样本 9
乡村型站 A	SI018	SI024	SI030	SI018	SI024	SI030	SI018	SI024	SI030
城镇型站 C	SI006	SI006	SI006	SI007	SI007	SI007	SI009	SI009	SI009
湿度 A>C 天数/d	87	90	90	80	90	70	87	89	78
湿度 A<C 天数/d	2	0	0	4	0	16	3	0	7
A>C 的占比/%	97	100	100	89	100	78	97	99	87

统计在表 3 中的数据反映,本地样本最低气温时段的相对湿度逐日对比结果,城市站点绝大多数情况下低于乡村站点。再次说明分型算法区分出的乡村站

点的湿度,在最低气温发生时段显著整体高于城市站点,乡村站点会有更多的霜、雾、露,分型运算符合客观存在。

表4 样本90 d数据里日气温极小值(T_{\min})城市站点与乡村站点的统计比较

	样本1	样本2	样本3	样本4	样本5	样本6	样本7	样本8	样本9
乡村型站 A	S1018	S1024	S1030	S1018	S1024	S1030	S1018	S1024	S1030
城镇型站 C	S1006	S1006	S1006	S1007	S1007	S1007	S1009	S1009	S1009
T_{\min} A>C 天数/d	0	0	0	14	15	20	1	1	5
T_{\min} A<C 天数/d	90	90	90	74	73	66	89	89	85
C>A 的占比/%	100	100	100	82	81	73	99	99	94

表4的数据反映出,绝大多数情况下,分型归类为城市站点(C)的样本站点的日气温极小值要高于归类为乡村型的站点,这种结果符合常识,再次说明分型算法准确有效。

特别说明的是,在这组随机抽取的6个数据样本站点中,3个城市型站点间距10~15 km;3个乡村型站点中,S1018距3个城市型站点间距20~30 km,距另外两个乡村型站点间距为55 km和65 km。

3 结束语

提出一种区域气象站自动分型的方法,分型算法能够综合反映站点下垫面差异,算法输出的结果在气象观测台站管理应用、结论的物理机制解析以及相关理论的建立完善等方面都有实质意义。方法采用的观测数据为一个空间区域内的气温差异量,是很重要的热动力基础值,但这个量的演化规律在以往台站管理中很少被发掘应用,根据算法输出结论的精细程度,利用这种方法可以解决的问题包括:对大量区域自动站的观测环境异常变动的自动化在线识别;对站点的布局进行客观研判和优化;与测量数据配套的精细化地形系数的自动生成;分布式观测系统的构建;以高精度监控阈值改善数据质量等。其中分布式观测系统架构下站点的互补替代与数据的平行应用较为常见,如解决成都站数据的替代与延续需求。

参考文献:

[1] 李林,孙赫敏,乔晓燕,等. 探测环境对北京气象站气温观测值的影响分析[J]. 气象与环境学报,2019(4):120-126.

[2] 徐新良. 基于气象站观测数据的京津冀城市扩展对气温变化的影响研究[J]. 地球信息科学学报,2019(7):1086-1096.

[3] 罗丽,郭建侠,陈一枝,等. 公路路面热效应对环境气温的影响分析[J]. 气象科技,2018(2):403-411.

[4] 徐园园,任永建,许存华,等. 城市化对咸宁地区气温变化的贡献分析[J]. 江西农业学报. 2017(5):101-106.

[5] 岳艳霞,智利辉,王磊,等. 石家庄市区域自动气象站气温数据适用性分析[J]. 气象与环境学报,2017(2):53-62.

[6] 张媛,任国玉. 无参考序列条件下地面气温观测资料城市化偏差订正方法:以北京站为[J]. 地球物理学报,2014(7):2197-2207.

[7] 李娇,任国玉,任玉玉,等. 资料均一化对沈阳站气温趋势和城市化偏差分析的影响[J]. 大气科学学报,2014(3):297-303.

[8] 茆佳佳,沈雪峰,郭建侠,等. 水体对气温观测影响的试验分析[J]. 气象科技, 2014(3):496-502.

[9] 沈滢洁,王成刚,曹乐,等. 屋顶绿化对城市降温效应的模拟分析——以南京市为例[J]. 气象, 2017(5):610-619.

[10] 贾红莉. 不同下垫面对气温观测的影响分析[J]. 南方农业,2016,10(9):220-221.

[11] 玉坤. 空间数据聚类在气象灾害预警中的应用[J]. 电子技术与软件工程,2018(9):164-165.

[12] 张瑶,历超,王子洋. 聚类分析在气象数据处理中的应用[J]. 林业勘查设计,2010(3):37-39.

[13] 严中伟,李珍,夏江江. 气候序列的均一化——定量评估气候变化的基础[J]. 中国科学:地球科学,2014,44(10):2101-2111.

[14] 贺南,文斌,陈乐,等. 一种新的气象数据分析和误差诊断方法[J]. 成都信息工程大学学报, 2019,34(4):342-345.

[15] 张学文. 熵气象学[M]. 北京:气象出版社, 1992:153.

[16] 屠其璞,翁笃鸣,武全,等. 气候资料超短序列订正方法讨论(上)[J]. 南京气象学院学报, 1978(1):59-67.

[17] 李庆祥,刘小宁,张洪政,等. 定点观测气候序列的均一性研究[J]. 气象科技,2003(1):3-10.

[18] 周子康. 山区气温超短序列延长方法的讨论[J]. 农业气象,1983(2):53-61.

[19] 吴雪. 城市化进程对气温变化的分析——以北京、郑州、南京、杭州城市为例[J]. 内蒙古气象,2019(3):42-45.

[20] 汤绪. 超大城市气象观测网的设计与应用, [J]. 气象科技进展,2016(4):8-17.

[21] 贺南,夏昕,杨乐,等. 一种区域自动气象站客观分群方法[P]. 发明专利,ZL2020.1.1253654.7.

Design and Application of Automatic Classification

Method for Chengdu Regional Weather Station

XIA Xin¹, WANG Dongmeng², HE Nan³

(1. Chengdu Meteorological Office Chengdu China Postcode, Chengdu 611130, China; 2. Chengdu University of Information Technology Chengdu China Postcode, Chengdu 610225, China; 3. Wenjiang Meteorological Office Chengdu China Postcode, Chengdu 610225, China)

Abstract: In order to quickly classify different types of weather stations on the underlying surface, an automatic classification method of regional weather stations is proposed. First, the temperature data selected from the stations where the difference between the two minimum daily temperature values is between $-5\text{ }^{\circ}\text{C}$ and $+5\text{ }^{\circ}\text{C}$, is divided with a step of $0.1\text{ }^{\circ}\text{C}$ as, and sorted according to the temperature difference value; then the frequency analysis sequence reflecting the sample size of a certain temperature difference position can be obtained from the temperature difference values of each interval within 90 days. And the median frequency of the normalized sample size is taken as the characteristic frequency, marked as F ; finally, the quantization factor is set K_f quantifies the factors affecting distance, and the influence of the underlying surface consistency can be reflected in K_f . The final results show that, the stations are automatically divided into 3 groups: A group, B group, and C group. By analyzing groups A and C, it can be concluded that the geographical distribution of rural and urban stations is clearly defined. The validity of the operational classification was verified by sampling inspection of group A and group C and comparing the data of urban and rural sites.

Keywords: classification; daily minimum temperature difference; frequency; underlying surface type